2017

The Rational Investigation of Anti-Cancer Peptide Specificity using the Knob-Socket Model

Shivarni Patel
University of the Pacific, s_patel18@u.pacific.edu

THE RATIONAL INVESTIGATION OF ANTI-CANCER PEPTIDE SPECIFICITY USING THE KNOB-SOCKET MODEL

by

Shivarni B. Patel

A Thesis Submitted to the

Graduate School

In Partial Fulfillment of the

Requirements for the Degree of

MASTER OF SCIENCE

Pharmaceutical and Chemical Sciences
Graduate Program

University of the Pacific
Stockton, CA

2017

THE RATIONAL INVESTIGATION OF ANTI-CANCER PEPTIDE SPECIFICITY USING THE KNOB-SOCKET MODEL

by

Shivarni B. Patel

APPROVED BY:

Thesis Advisor: Jerry Tsai, Ph.D

Committee Member: Liang Xue, Ph.D.

Committee Member: Geoff Lin-Cereghino, Ph.D.

Department Chair: Andreas Franz, Ph.D.

Dean of Graduate School: Thomas H. Naehr, Ph.D.

.

THE RATIONAL INVESTIGATION OF ANTI-CANCER PEPTIDE SPECIFICITY USING THE KNOB-SOCKET MODEL

by

Shivarni B. Patel

DEDICATION

This thesis is dedicated to my parents for taking my call at 2 o'clock in the morning because I didn't check the time difference; my brother, Nikash, for helping me keep my title as *Favorite Child*; my grandparents for making sure I am never without love or dal bhaat; my fiancé, Brandon, for putting up with a ridiculous number of mood swings; and of course, Teya for very easily becoming my favorite #DisneyBuddy4Life.

ACKNOWLEDGMENTS

My gratitude goes to Dr. Jerry Tsai for his guidance during my time at Pacific. He has constantly been supportive and understanding as I struggled both in and out of school during my time as his graduate student. I would also like to thank Dr. Hyun Joo for patiently helping me navigate through every program that was used to put this paper together. Thank you to Keith and Ferdiemar- we came into the program together and without them I would not have made it through half of the classes or teaching semesters. Thank you to Taylor Rabara- your addition to the Tsai group has made my time in the program that much more enjoyable and I am grateful for the friendship I have gained. The work done for the PDZ domain and Bcl-2 proteins would not be possible without the help of Stephen Tang, Zaina Chaban, Nathaniel Chien, and Vivian Kellner. I would like to thank the University of the Pacific, Chemistry Department as well as the Pharmaceutical and Chemical Sciences Graduate Program for giving me the opportunity to continue my education. Finally, I would like to thank the members of my Master's thesis committee for their help in the revision of the work presented and the completion of this thesis.

The Rational Investigation of Anti-Cancer Peptide Specificity using the Knob-Socket Model

Abstract

by Shivarni B. Patel

University of the Pacific
2017

Cancer has been a pervasive and deadly problem for many years. No treatments have been developed that effectively destroy cancer cells while also keeping healthy cells safe. In this work, the knob-socket construct is used to analyze two systems involved in cancer pathways, the PDZ domain and the Bcl-BH3 complex. Application of the knob-socket model in mapping the packing surface topology (PST) allows a direct analysis of the residue groups important for peptide specificity and affinity in both of these systems. PDZ domains are regulatory proteins that bind the C-terminus of peptides involved in the signaling pathway of cancer progression. The domain includes five β-strands, two α-helices, and six coils/turns. In this study, the PST of all eight solved crystal structures of T-cell lymphoma invasion and metastasis 1 (Tiam1) PDZ domains are mapped to reveal details of ligand-domain binding pockets and packing interactions. Four main interactions were identified in the comparison of the PST maps and a consensus sequence was calculated using knob-socket interaction data. In the case of the Bcl-BH3 complex, binding of these two proteins prevents an unhealthy cell from undergoing apoptosis. In

the knob-socket mapped protein-ligand interactions, the helical ligand consists of 8 to 10 residues that specifically interact with four helices on the binding protein: the N-terminus of Helix2, the main bodies of Helix3 and Helix4 and the C-terminus of Helix5. Among all of the interactions that were analyzed, there were three amino acids from the ligand, glycine, leucine, and isoleucine, that always packed into the hydrophobic groove that is key for ligand recognition. By using knob-socket analysis to map quaternary packing structure, it was possible to identify the quaternary-level protein interactions that define ligand specificity and binding strength. From this analysis, possible protein mimetics can be developed that could be used as cancer treatments.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1: Introduction

For many years, cancer has been one of the greatest obstacles for mankind. Despite pouring billions of dollars of funding into discovering a cure, there is yet to be a definitive cure. Recent research, however, has provided a new potential cancer treatment that may become the key to finding an effective treatment. The body's immune system normally works by destroying any harmful tissues through the process of apoptosis, or programmed cell death. In apoptosis, unhealthy cells will destroy themselves in order to prevent their reproduction (Petros, Olejniczak, & Fesik, 2004). Although apoptosis is a normal cellular process, there are many checks and balances that regulate the process so that it does not exceed what is necessary and kill too much normal tissue. However, cancers take advantage of these "shut off switches" in the body to dampen immune response and help them survive. Studies have been conducted that show that there are certain molecules that dampen the immune response that cancer cells take advantage of, and it has been shown that by temporarily disabling these molecular "brakes" with synthetic proteins, the immune system can attack cancer cells more effectively.

The PDZ domain is involved in a signaling pathway that regulates cancer progression (H. J. Lee & Zheng, 2010). Its structure comprises of a five-stranded β-sheet sandwiched between two α-helices (S. O. Lee et al., 2016). The specific ligand that binds the domain depends on the family from which the domain belongs. In general, the

domain will bind the C-terminus of the ligand (Shepherd & Fuentes, 2011; Shepherd, Hard, Murray, Pei, & Fuentes, 2011). Many research groups have agreed that the interaction between the ligand and the PDZ domain involves the second strand, the second coil, and the second helix of the domain's sequence (Ernst et al., 2014). The analysis conducted in this work shows that not only are there side-chain interactions between the ligand and those three pieces of the PDZ domain, but the ligand goes so far as to extend the domain's β-sheet. By analyzing the domain-ligand binding structure, there is a possibility of intervening in the binding of the PDZ domain and its ligand.

The interaction between members of the Bcl-2 and BH3 families have also been linked to cancer progression (DeBartolo, Dutta, Reich, & Keating, 2012). By binding to one another, the resulting protein complex halts a cell's ability to undergo apoptosis and therefore allows for cancerous cells to survive and multiply (Delbridge & Strasser, 2015). The binding ligand is helical as is the protein to which it binds, which consists of seven helices connected together by short coil/turn segments (Dutta et al., 2010). A two-dimensional representation of this helical complex would make it easier to identify patterns that could be essential in future protein design and binding studies. A designed protein, or in this case a BH3 mimetic, would allow cancerous and unhealthy cells to undergo apoptosis. So, a better understanding of this protein-protein interaction is a key to better cures for diseases such as cancer.

The knob-socket construct presents a way to better understand protein packing interactions at the quaternary level (Joo, Chavan, Phan, Day, & Tsai, 2012). It is a tetrahedral motif that identifies a three-residue socket from a single protein secondary structure which packs a single-residue knob from another secondary structure in order to define tertiary and quaternary packing. This model uses data collected from solved structures in the Protein Data Bank (PDB) in order to make mapping of these proteins on a two-dimensional map possible. By displaying a protein's structure on a single plane, protein-protein interactions not only become more apparent, but patterns among members of the same protein family can be seen as well. The use of the knob-socket model allows for the identification of discrete units of interactions between two proteins.

Overall, the knob-socket model provides a mapping interface that allows the ease of visibility of patterns and common sequences that would be essential in future protein binding studies. By isolating key components of the PDZ domain-ligand interaction and the Bcl-2 and BH3 complex, peptides of higher specificity could outcompete the current binding ligands that are preventing a cell from halting the progression of cancer. With a model in place for both of these systems and the verification of knob-socket observed patterns with the work of other research groups, further steps could be made to produce more specific and effective cancer treatments.

Chapter 2: PDZ Domain Peptide Binding Design by use of the Knob-Socket Model

PDZ stands for PSD-95 (postsynaptic density PSD-95/SAP90) DLG (Drosophila melanogaster tumor suppressor septate junction protein Disks large-1) ZO1 (epithelial tight junction protein Zonula Occludens-1) (2016). This domain is present in many different proteins, but its general structure remains the same with two helices on both ends and a 5-stranded β-sheet in the middle (Figure 1C). The PDZ domain acts as a regulatory center, interacting and binding with members of a signaling pathway (Smith & Kortemme, 2010; Smith et al., 2013). Regulation of a signaling pathway by the PDZ domain is done by binding the C-terminus of a protein (Ernst et al., 2014; Kundu & Backofen, 2014; Mignon, Panel, Chen, Fuentes, & Simonson, 2017).

When looking specifically at the peptide binding into the PDZ domain, the function of the domain becomes apparent. Syndecan1 (SDC1) is a tumor suppressor that has been associated with breast cancer, lung cancer, and mesothelioma (Cheng et al., 2016). By recognizing and binding to the PDZ domain, SDC1 is able to prevent the detrimental characteristics of cancer cells, such as increased cellular adhesion, migration and resistance to irradiation (Cheng et al., 2016). One of the problems associated with SDC1 is the inverse relationship it has with the progression of cancer in the body. If SDC1 is decreasing in expression levels as the cancer cell metastasis rate increases, the positive function of SDC1 in the presence of cancer cells becomes insignificant.

Therefore, a more descriptive model of the specific determinants of domain-peptide interaction would aid in the advancement of controlling cell signaling pathways (Ernst et al., 2014).

In this study, we took all eight solved structures of the Tiam1 PDZ domain from the PDB and mapped them using the knob-socket model. The map highlights the similarities and patterns in the binding motif. The ability to identify these patterns would aid in the *de novo* design of SDC1-mimics that could artificially act as SDC1 would in the presence of cancer cells.

The C-terminus binding scheme of a polypeptide to the PDZ domain can be seen in Figure 1C, which shows the syndecan1 peptide and Tiam1 PDZ domain interaction. Tiam1 (T-cell lymphoma invasion and metastasis 1) functions as a guanine nucleotide exchange factor (Liu, Shepherd, Murray, Xu, & Fuentes, 2013). The typical bound peptide of the PDZ domain in this protein is eight residues in length. The naming of the residues involved in binding starts with the last, C-terminus residue on the peptide, identified as P_0, and works backwards with next three residues labeled as P_{-1}, P_{-2}, and P_{-3} up to eight residues, or P_{-7}.

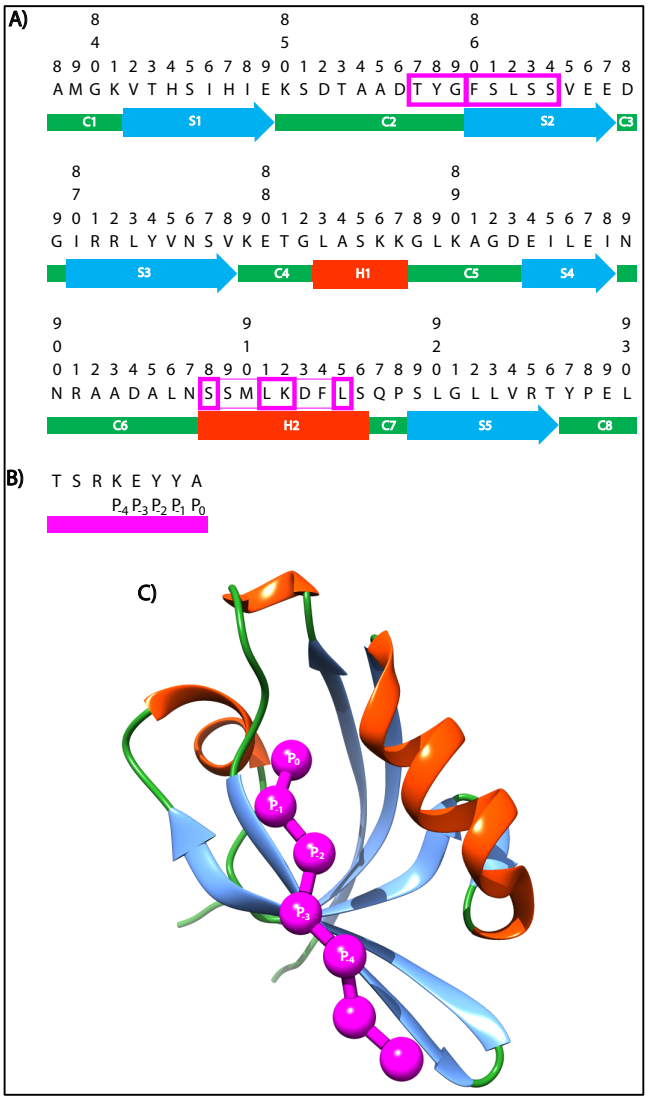


Figure 1: PDZ Domain Sequence and Structure. (A)The sequence of a Tiam1 family PDZ domain (PDB ID: 4GVD_BC) is shown, with its secondary structure indicated underneath. Green lines represent random coil, blue arrows represent strands, and red lines represent helices. Each secondary structure has been labeled in white text to indicate its location in the sequence. Pink boxes around the sequence indicate locations in which a bound peptide has been shown to interact with the domain, as defined by the knob-socket model. (B) The sequence shown is that of the bound peptide for this domain, with appropriate labeling of the residues shown underneath. (C) The three-dimensional structure of the PDZ domain with a bound peptide, shown in pink. The image shown was created using the UCSF Chimera application.

The binding preference and specificities of the PDZ domain and its family members are only beginning to be understood. A recent study on syndecans shows that they have an EFYA sequence that binds to PDZ domain-containing proteins (Cheng et al., 2016). Specifically for the PDZ domain of Tiam1, the interaction with a syndecan regulates cell adhesion to fibronectin (Cheng et al., 2016). Many other studies have been done to improve the knowledge of the PDZ domain and its binding motif. Since finding a pattern in the binding motif is the main goal for many current studies, it is important to take note of mutational work on the domain and its bound peptide to identify specificity levels. One study used phage-displayed random peptide libraries to identify binding specificities of the PDZ domain (Tonikian et al., 2008). Ultimately what was found was a versatility to the PDZ domain and its ability to adapt to mutational studies conducted on the domain itself. The wild-type domain has a tendency to bind hydrophobic C-termini peptides. When mutations were added to various residues on the PDZ domain, there was one pattern identified that could help with progressing the knowledge of the PDZ domain binding specificity. This pattern involved the residue on the peptide in position P_{-2}. The specificity for this position is brought about from interactions with the PDZ domain's second helix, and the sequence at the interaction site on the helix plays a vital role in the binding of the P_{-2} residue. By identifying the helix as an important interaction site for the peptide as well as determining sequence specificity in the PDZ domain, we would be able to enhance any attempted binding studies with lab-made peptides. Another binding study, which related closely to the work in this paper, aimed to identify binding patterns

between the Tiam1 PDZ domain and the syndecan family (Liu et al., 2013). This study concluded that along with a binding preference for syndecans -1 and -3, the PDZ domain also had two binding pockets that interacted with syndecan1. The first pocket, named S_0 by this group (not to be confused with nomenclature used by the knob-socket model later) was formed by residues Y858, F860, and L915. This particular pocket bound the P_0 residue of the syndecan1 peptide. The second pocket, named S_{-2}, was formed by the PDZ domain residues L911 and K912 to interact with the P_{-2} residue of the peptide. With both of the described studies, a conclusion has been made regarding the correlation between the sequence patterns and the binding structure. Both of the conclusions, though slightly different, can be used to develop a better understanding of the common binding scheme used in the PDZ domain with its bound peptides.

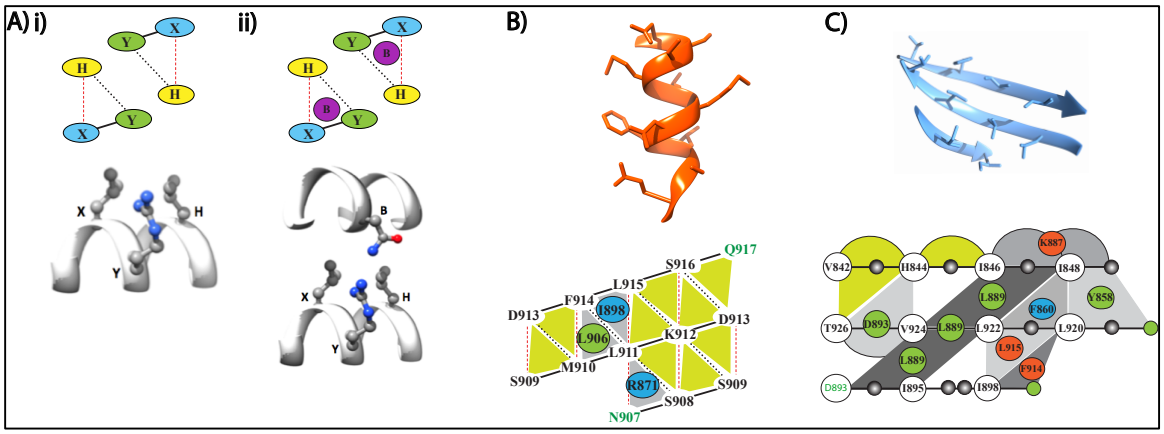


Figure 2: Knob-Socket Binding Motif. (Ai) Two-dimensional representation of a free socket. Three residues make up each socket: X, Y, and H. Residues X and Y are next to each other in sequence and therefore share a peptide bond, indicated by a solid black line. Residues X and H share a hydrogen bond, indicated by a red dotted line. Residues Y and H share a Van der Waals interactions, indicated by the black dotted line. (Aii) Two-dimensional representation of a filled socket. Residue B is shown in a purple circle and is a residue from another secondary structure, therefore indicating tertiary or quaternary folding. This residue interacts with all three amino acids of the socket through side-chain interactions. (B) Modified lattice for an α-helix used to visualize protein folding as defined by the knob-socket model. A portion of the helix has been mapped out, with free socket colored in green, and filled socket colored in grey with the appropriate knobs inside. (C) Map used to visualize a three- stranded β-sheet. Since sheets are made up of different strands connected by coils of varying length, the map for a sheet is very different from that of a helix. The same visualization can still be achieved by this map, though: Knobs from other secondary structures are shown in red (helix), green (coil), and blue (sheet) circles and free sockets are still indicated in green and filled sockets in grey.

There is still a need to further analyze both the common sequence and the structure of the PDZ domain to aid in identifying a motif for the binding peptide. In this work, we will discuss the use of the knob-socket model in visualizing the binding pattern of the PDZ domain in Tiam1 and the conclusions the model allows us to make regarding a consensus sequence for a binding peptide. The knob-socket model is the packing motif used in this study for determining the interaction structure of the PDZ domain in Tiam1. As Figure 2A depicts, this tetrahedral motif identifies a three-residue socket from one secondary structure and a side-chain knob from a residue of another secondary structure. This knob B residue packs into the socket, XY:H, forming the 3^{o} and 4^{o} structure of a protein. The propensity of a socket to be filled or free of a knob is determined by the residues comprising of each individual socket (Joo et al., 2012). A modified lattice is used to show the knob-socket motif on a two-dimensional surface as a packing surface topology (PST). By simplifying the structure of proteins onto these lattices, we are able to identify possible packing surface patterns. Generally, the three residues of a socket are labelled X, Y, and H in which X and Y are sequential, connected by a peptide bond, X and H are connected by hydrogen bonds, and Y and H are connected by van der Waals interactions. As Figure 2C and 2D shows, helices and sheets are represented differently. In helices, the modified lattice represents the described bonds by placing X in position i, Y in position i+1, and H in position i+4. Any combination of amino acids in these positions relative to X form the triangular sockets. Sheets are labelled differently, mostly because the number of amino acids between residue X and residue H is unique to each

protein. It is also known that in sheets, the alternating side-chains are facing opposite directions. This would mean that sockets can be formed on the "front side" and "back side" of the sheet, with residues X and Y no longer being consecutive. For this reason, X is identified as being in position i, Y is either in position i+1 or i+2 due to alternating side-chain positions, and residue H is in position j. With this model, advancements in drug design and the inhibition of binding of cancer-causing regulatory peptides to the domain could be made.

The use of the knob-socket model to map the PDZ domains of Tiam1 aid in the identification of patterns among the sequences of the bound peptides as well as similarities in the packing interactions between the PDZ domain and the peptide on a two-dimensional level. The sequence information for each of the eight available Tiam1 PDZ domains were collected from the PDB and run through secondary structure prediction systems and then analyzed using the knob-socket motif. With the data from both of these processes, a PST for each of the PDZ domains were uniformly made so differences and similarities could easily be identified. Three of the eight structures had information with a bound peptide in the domain which was useful in making conclusions regarding the packing availability of the PDZ domain. For the Tiam1 PDZ domain PST shown in Figure 1, the main interaction sites involve an extension of the β-sheet as well as knobs from the binding peptide on the backside of the second coil and in the second helix of the domain itself. Therefore, the knob-socket analysis produces a model to

explain the binding of C-terminal peptides based on a set of definable and dependent packing interactions.

Materials and Methods

PDZ domain data collection. All solved structures of the PDZ domain were found using the RCSB Protein Data Bank (PDB) (Berman et al., 2000). The PDB is an open-access library of all solved three-dimensional structures of various biomolecules, which are primarily proteins. These molecules are organized by their assigned four-character PDB ID and each contains references to the literature responsible for aiding in solving the structure and providing the three-dimensional analysis. Using the PDB, 162 structures coming from 29 different PDZ domain families were found. The PDB files associated with each of these structures contain primary and secondary structure data, which were then visualized in three-dimensional form using the program, Chimera.

Molecular representation and visualization. UCSF Chimera is a molecular modeling program that can be used for three-dimensional visualization of proteins by pulling data from the PDB (Pettersen et al., 2004). Figure 1C shows that the structure of a Tiam1 PDZ domain and its bound peptide as it was rendered in Chimera. The coloring scheme and orientation of the structure were kept uniform through this work on the PDZ domain for consistency. For each image, coils were colored forest green, helices were red-orange, strands from β-sheets were cornflower blue, and the bound peptides were set to have a magenta color. When it came to orientation of each domain, the bound peptide

was kept in the center, the second helix of the PDZ domain was to the right of the peptide while the last strand of the domain's five-stranded β-sheet was on the left, and the second coil of the domain was above the peptide. This orientation was chosen so that all interactions of the domain with a bound peptide can easily be seen without the use of interactive programs.

Knob-socket protein surface topology (PST) map of packing structure. The knob-socket model was the packing motif used to better understand the interactions involved with the PDZ domain. Using this model provided a clear definition of secondary, tertiary, and quaternary structure (Joo et al., 2012). Sockets are formed by a combination of three residues in a single domain, usually close in proximity when the protein is folded, forming primary and secondary structure. Knobs are single residues that originate from either a different part of the same domain or from a different domain altogether, and can therefore be responsible for the tertiary and quaternary structure. To better understand how knob-socket packing works, it is best to look at the two-dimensional maps used to represent higher order protein folding.

Helices were mapped using a unique, modified lattice shown in Figure 2A and 2B. In this figure, sockets are represented by each triangle, formed by three residues. It is brought together by a peptide bond between consecutive residues (solid black line), a hydrogen bond between every four residues (red dotted line), and a van der Waals interaction between every three residues (black dotted line). The position nomenclature

for the residues in a helical socket can either be i, i+1, and i+4 or i, i+3, and i+4 positions. The free sockets shown in Figure 2Ai best represents both of these combinations using residue labels X, Y, and H. The "low" X socket on the left has X in the i position with Y and H in the i+1 and i+4 positions, respectively. The "high" X socket on the right has H in the i position with X and Y in the i+3 an i+4 positions, respectively. The image underneath these two sockets shows the proximity of residues X, Y, and H on a helix.

The knob-socket model uses the Voronoi Polyhedra/Delauney Tessellations to define contacts and packing order in the different secondary structures (Dutour Sikiric, Garber, Schurmann, & Waldmann, 2016). Each individual socket has a specific propensity to pack a knob. As discussed in previous literature (Joo et al., 2012), socket propensities are the frequency in which sockets are found to be free or filled of a knob. It also defines the tendency for a socket to form at all, with non-forming sockets being named non-sockets. By using all solved protein structures in the SCOP database, frequencies can be built, directly relating to the tendency for any three-residue combination to form filled or free sockets or non-sockets.

Knobs tend to pack into sockets of higher packing propensities. Figure 2Aii shows a packed knob B into socket XY:H. A filled lattice is shown in Figure 2B. The colors in this map represent different types of sockets and knobs coming from different secondary structures. Grey sockets represent filled sockets with a packed knob. Many times, filled sockets located close to one another are colored in with different shades of

grey. This variation in the shading is solely for the purposes of distinguishing different sockets, and does not represent any propensity data. Green sockets are free sockets. Knobs are colored based on the secondary structure they come from- green knobs come from a coil, red knobs come from a helix, and blue knobs come from a sheet. For the purposes of the research in this paper, the knobs that come from a bound peptide in the PDZ domain are colored a magenta color.

Sheet packing has its own mapping system through the knob-socket model as well. The residues on a single strand of a sheet are alternating in direction and residues that line up together from one strand to the next are facing the same direction. A filled map for a sheet is shown in Figure 2C. Residues facing out of the strands are labelled with their single letter amino acid code and residue number in hallow circles. Residues facing the opposite direction, away from the viewing position, are represented by small grey circles. With sheet mapping, the front and backside of the sheet both need to be represented for a full understanding of all the interactions. Since the number of residues between the strands of a sheet is unknown, the nomenclature used to describe the residues of a socket in a sheet is different. The first residue of a socket is labelled as residue i, with the next residue being two amino acids away in the i+2 position, and the third residue in the j position. All three residues in a socket are facing the same direction. The coloring of sockets and knobs are the same as helices, with a higher likelihood of finding different shades of grey for the sockets due to the proximity of the sockets in a sheet with separation by a type of bond like there is in a helix map.

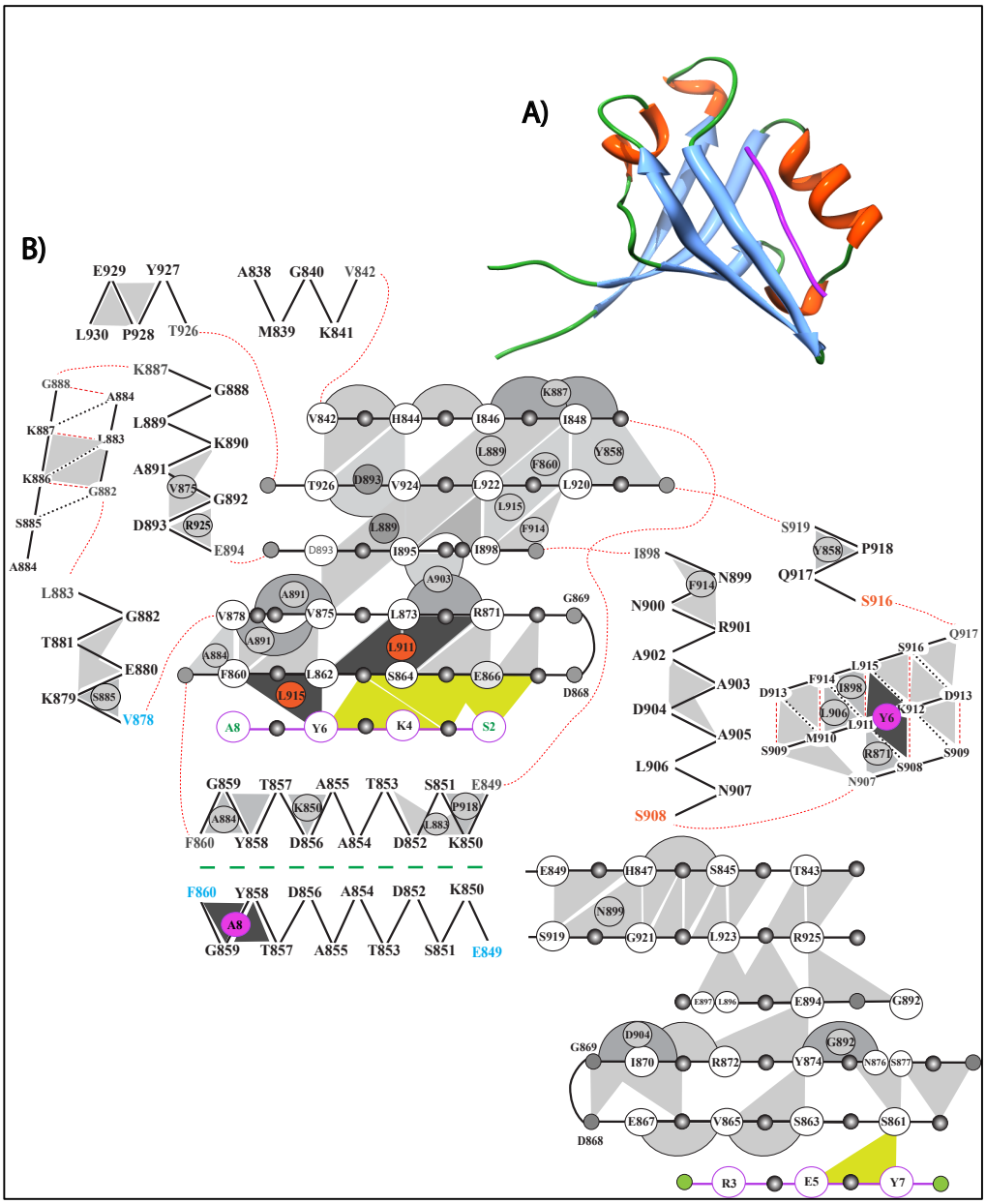


Figure 3: Full Knob-Socket Map for Bound PDZ domain. The PDZ domain shown (PDB ID: 3KZE_BD) has been fully mapped out using the knob-socket model. All except the bound peptide interaction have been greyed out to bring focus to the interaction sites. It can be seen that the bound peptide contributes knobs to the helix and coil of the domain as well as extends the sheet through sidechain interactions. The helix of the domain also contributes knobs to the domain's sheet when a bound peptide brings those two structures closer in proximity to one another.

Adobe Illustrator was used to create all of the maps according to the knob-socket model. Its features allow for a simple creation of all structural maps and the ability to create uniform maps for all of the domains analyzed. Figure 3 shows a completed map for one of the PDZ domains. The coloring was altered to accentuate the bound peptide, but the main features of the maps are still apparent. The red dotted lines connect the different secondary structures of the full domain together, with the map starting at the coil at the top of the image. The sheet is arranged in the center so that all residues connecting each of the strands are able to fit on the outside of the image. Lastly, the backside of the sheet is mapped to the side of each image. Together, these maps can be viewed side-by-side for a comprehensive analysis of each of the PDZ domains and their bound peptides.

Model of PDZ determinants of peptide ligand binding. The PDB data for each domain was collected and the knob-socket data was calculated. Each of the domains from the Tiam1 family was mapped out. Based on the key interaction sites, a general model shown in Figure 4A. The first of these interactions is between the final residue of the bound peptide, labeled residue P_0, and Coil2 of the PDZ domain. The residue P_0 packs into a pocket, a combination of multiple sockets next to one another, created by the first residue of Strand2 in the domain, B_0, and the final three residues of the coil, named C_0, C_{-1}, and C_{-2}. The next three interactions have two parts to them, any free socket interactions to show sheet formation without packing as well as the packed interactions. For the second interaction on the chart, the first and third residues from Strand2, B_0 and

B_2, form a socket with the last residue of the peptide, P_0, to pack the knob H_0 from Helix2 in the domain. The third interaction involves the residue P_0 and P_{-2} from the peptide forming a socket with Strand2 residue B_2 to pack the knob H_{-3} from Helix2. The fourth and final interaction in the sheet of the PDZ domain is a pocket formed by Strand2 residues B_2 and B_4 with peptide residues P_{-2} and P_{-4}. This pocket packs the Helix2 residue H_{-7}. The fifth interaction involves a pocket in Helix2 of the domain, formed by residues H_0, H_{-3}, H_{-4}, and H_{-7}. This pocket is packed with knob P_{-2} from the bound peptide. All of these interactions have been isolated and placed in a table, shown in Figure 4B.

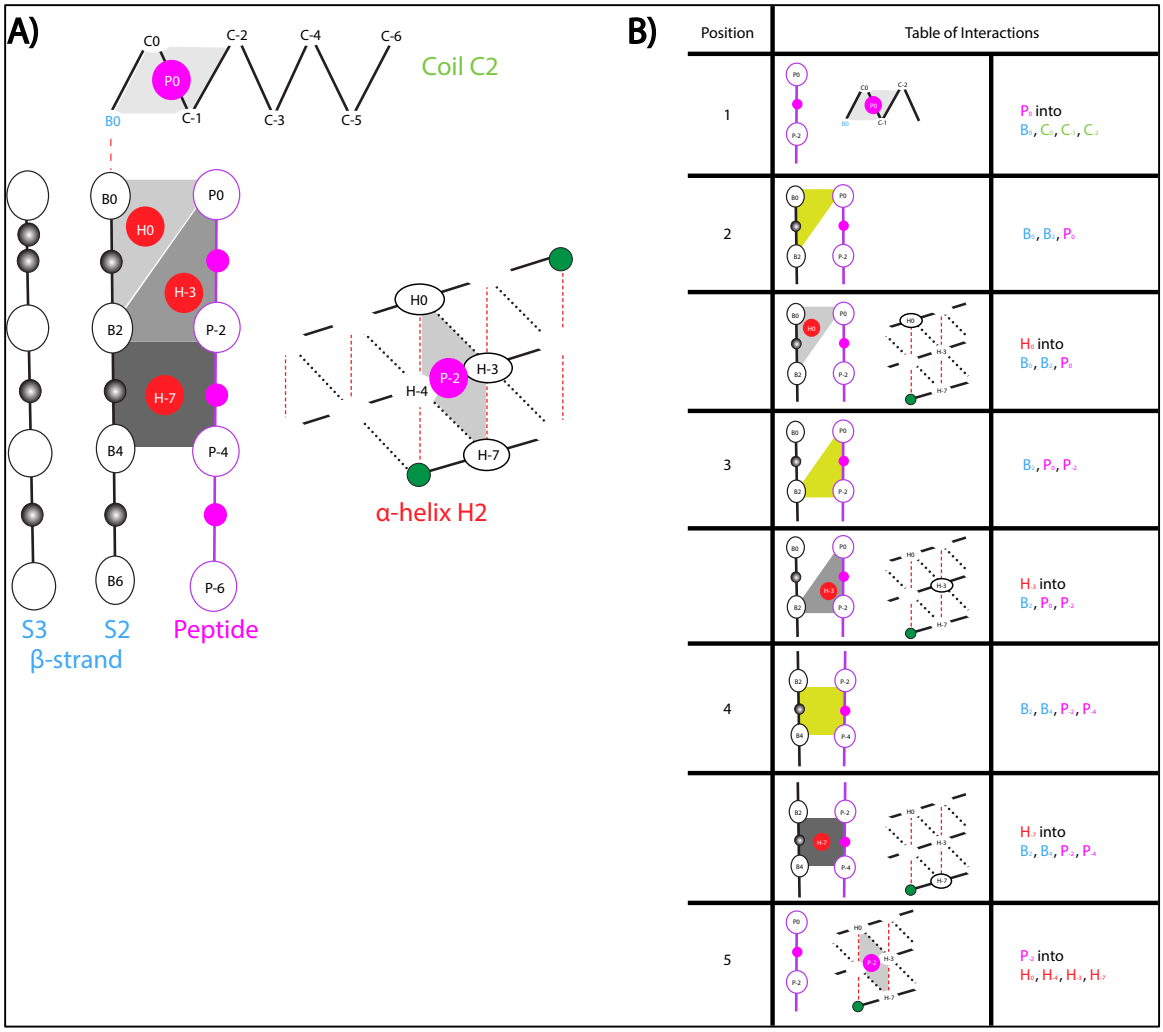


Figure 4: PDZ Domain Peptide Binding Model. (A) General map of the domain-peptide interactions seen most frequently using the knob-socket model. The bound peptide (pink) is shown to contribute knobs to Coil2 and Helix2 of the PDZ domain as well as extend the sheet at Strand2. It is also seen that Helix2 contributes knobs to the newly extended sheet. Each of the residues in the domain have been given generic labels so that this model could be applied to other PDZ domains and families. (B) Each of the interactions seen when a peptide binds to the PDZ domain have been isolated in this table. Each interaction between the newly extended sheet and Helix2 can be further separated into free and filled interactions, as seen in Positions 2, 3, and 4.

Optimal peptide ligand based on knob-socket analysis. With a general model in place, any PDZ domain with a bound peptide can be organized to determine not only how well the current peptide fits but also if there is a possibility for a better-fitting peptide for the domain. These peptides can be determined using the model based on the knob-socket analysis. For each socket and pocket, there is a specific propensity in which certain residues will pack into them (Joo et al., 2012). By analyzing these propensities for the Tiam1 family, a peptide with potentially better binding specificity could be determined. First, each of the previously described interaction sites from Figure 4B were combined into the following four categories: (1) extension of the domain sheet at Strand2, (2) packing of Helix2 knobs into the newly extended sheet, (3) knobs from the peptide that bound into Helix2 of the domain, and (4) knobs from the peptide that bound into Coil2 of the domain. Based on the known amino acids of the PDZ domain in the Tiam1 family, the highest propensity sets of residues that satisfy each of the four interactions were plotted in Figure 5. A consensus sequence can be determined by observing common residues of high propensities in similar bound peptide positions, and a logos plot was determined for each of these four separate interactions. Unlike previous logos plots, positions in the peptide ligand are interrelated based on the PDZ binding model described above.

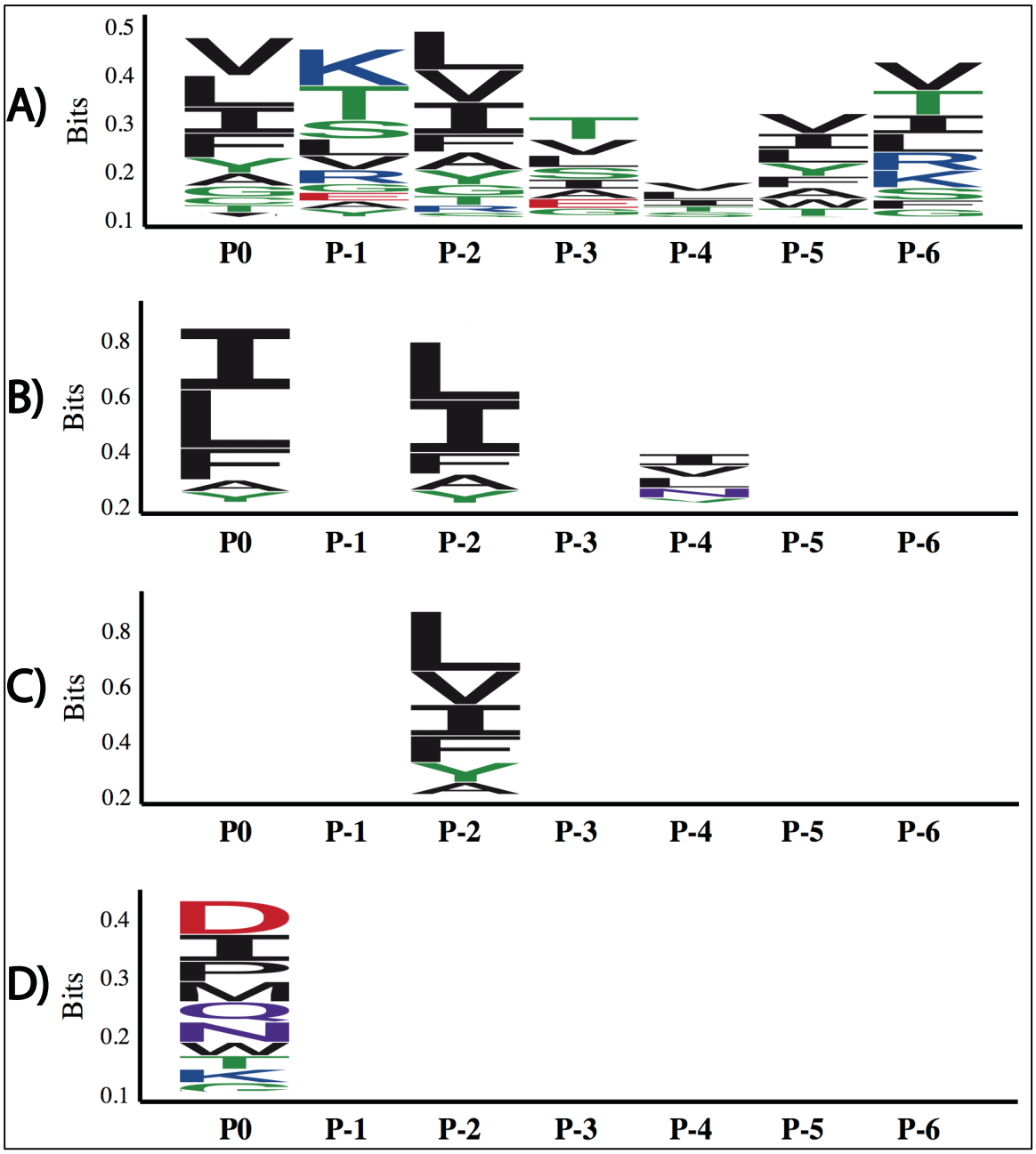


Figure 5: Logos Plots for Optimized Bound Peptide. (A) Plots the best peptide sequence to extend the sheet and therefore create free pockets between the domain's Strand2 and this bound peptide. (B) Plots the best peptide residues in positions P0, P-2, P-4 to allow the domain's Helix2 knobs to pack into the newly extended sheet. (C) Plots the best peptide knob in position P-2 to pack into Helix2 of the domain. (D) Plots the best peptide knob in position P0 to pack into Coil2 of the domain.

Results and Discussion

Quaternary packing analysis using the knob-socket motif. By providing a functional construct to describe residue packing in proteins, the knob-socket model can provide a detailed and discrete assessment of a binding interface. In the instance of the PDZ domain, there are three main areas of contributions that were common among a majority of the domain families. The most significant of these interactions was the extension of the domain's sheet with the bound peptide. Not only are filled sockets created by the extension, but significant hydrogen bond mediated packing interactions between strand residues are formed. The presence of both filled and free sockets are equally as important when considering binding interactions because they show that proximity of the residues has allowed for both hydrogen bonds and Van der Waals interactions to occur. In the PDZ domain, ligand binding creates new sockets that extend the sheet beyond just creating point interactions with the bound peptide. For example, there is a domain that did not have any packed interactions in the sockets that are found between Strand2 of the PDZ domain and the bound peptide. There were, however, free sockets between the domain's sheet and the bound peptide, showing that the sheet was holding the peptide close. The importance of the free sockets does not take away from the filled sockets that form, though. The extension of the sheet using the bound peptide provides a place for residues from Helix2 of the domain to pack into the sheet. This interaction would not be possible, and is not even seen, when there is no bound peptide.

Without the bound peptide, the helix is not close enough to the sheet without the peptide to find any pockets that would accept its residue knobs.

Before going into the specifics of the domain-peptide interactions, it is important to first define a labelling scheme. This labelling scheme can be seen in Figure 4A. Here, all significant residues for this study have been changed so that the model can be applied to any of the domains once mapped out using the knob-socket model. The residue labels use subscripts to reflect their position relative to one another on the piece of secondary structure they are representing. For Coil2 of the domain, C_0 is the first residue of that secondary structure with a pocket being formed starting from 2 residues away and ending with the B_0, which is the first residue of Strand2 which is next in sequence in the domain. In Strand2, the first residue of the strand is B_0 with all residues following it increasing in numerical value (ex. 0, 2, 4, …). Finally, for Helix2 the residue labelling starts at the end of the helix with residue H_0, and decreases numerically from there (ex. 0, -1, -2, …). The labelling for the bound peptide was pulled from previous literature (Fuentes, Der, & Lee, 2004), with the C-terminus residue labelled P_0, and all residues prior to it decreasing in numerical value.

With this labelling system in place for referring to the different residues in each bound PDZ domain map, the interactions can be looked at and compared to one another. The first major interaction occurs between Strand2 of the domain and the peptide. There are two sockets and one pocket that formed in this newly extended sheet space. The first

of these sockets was made up of residues B_0, B_2, and P_0, when labelled according to the scheme used in this study as seen in Figure 4A. Looking at Figure 3, these residues are F860, L862, and A8, respectively. The second socket is made up of residues P_0, P_{-2}, and B_2, which is A8, F6, and L862, respectively in the figure. The final free pocket found in this space is between residues P_{-2}, $P_{-4,}$ B_2, and B_4. These residues are F6, E4, L862, and S864, respectively, in Figure 3.

While these three pockets were used in creating the model for the bound PDZ domain structure, it is important to note that not all of the domains contain each of these interactions, but instead some variation of them. Each pocket can be either filled or free as well as being a modified version of the model seen in Figure 4A. For the instance of the Tiam1 domain mapped in Figure 3, not all of these interactions are exactly located where the model places them, but when comparing each of the maps together from different families, the consensus model pulled out the most common sockets and pockets can be seen.

This concept applies to the packing pattern that was created for the model as well. Each of the sockets and pockets were predicted to pack one of three significant residues from Helix2 of the domain, H_0, H_{-3}, and H_{-7}. While there are variations to this packing, which can even be seen in the sample map in Figure 3, the consensus packing pattern for the model was chosen because it represents what was seen to be the strongest packing pattern amongst all of the maps made for the PDZ domain and its families.

The second and third interactions are secondary to that of the extended sheet. This is because if the sheet does not pack the peptide correctly, the peptide would not come close enough to the helix and maximize its strength in the domain. The second and third interactions from between the bound peptide and Coil2 and Helix2 of the domain. While it was just discussed that the helix contributes knobs to the sockets formed in the extended sheet, there is also the possibility of the peptide contributing knobs to the sockets formed in the helix as well as with the coil located above the final residue of the peptide. Looking back at Figure 3, it can be seen that the final residue of the peptide, P_0, binds into a four-residue pocket between the final three residues of Coil2 and the first residue of the following strand in the domain sequence. Similarly, the second to last residue in the peptide sequence, P_{-2}, binds into a pocket formed by four helix residues, labelled H_0, H_{-3}, H_{-4}, and H_{-7}.

Knob-socket defined binding model. Based on the data collected from the Tiam1 family, all three of these interactions were brought together to create a general model, as seen in Figure 4. Figure 4A has been used thus far to show the labelling method used for each of the residues in the model. Figure 4B is a table that has been created to separate all of the model's interactions that were identified. These interactions have not been organized by contribution to binding on the table, and it is important to note that positions 2, 3, and 4 possess two parts to their interactions. Because they are both domain sheet and peptide interaction sites, each of these sockets or pockets has a free and filled propensity, both holding equal value in a bound ligand domain's structure.

Position 1 binding involves Coil2 packing with P_0 of the peptide. Position 2 has two parts to its observed interactions. The first was the free socket formed between residues B_0, B_2, and P_0, showing a sheet extension. The second interaction is the packing of H_0 into the socket formed by the same residues. Position 3 exhibits the same two interaction types between residues P_0, P_{-2}, and B_2, packing H_{-3}. Position 4 also has the same two interaction types, forming a pocket between residues P_{-2}, $P_{-4,}$ B_2, and B_4, and packing H_{-7} from Helix2 of the domain. The final interaction on the table, Position 5, occurs between Helix2 with P_{-2} of the bound peptide.

This general model holds true across the Tiam1 family and a majority of members from other PDZ domain families. While there are differences in some domains, the order of importance of these interactions when binding a peptide remains consistent overall.

PDZ domain specificity. The difference between affinity and specificity is important when looking at PDZ domain binding. With a peptide ligand, specificity refers to the sequence of the peptide packing into the domain, while affinity determines how strong the binding interaction will be between the peptide and the domain. If the residues of the PDZ domain do not accommodate the packing of the sequence, then the peptide will not pack. But if the domain contains sockets and pockets that prefer packing, or have a high propensity to pack, then the peptide is more likely to pack in the domain. The concept of sockets having propensities to pack or not pack is determined by data collected from the PDB (Joo et al., 2012). With affirmation that sockets of similar

residue composition pack specific knobs, a socket propensity can be linked to the strength and confidence of higher order packing.

This study is looking to use the knob-socket model to identify a peptide of high specificity and affinity to the PDZ domain. By breaking down each of the known interactions as shown in Figure 4B, the different contributions to the binding could be assessed to find the best peptide sequence for the PDZ domain. The results of the data in Figure 5 will be discussed shortly. With an understanding of how the interactions between the PDZ domain and peptide are linked on a smaller scale, it is possible to make an informed prediction of a "better-binding" peptide for the Tiam1 family of the PDZ domain.

Knob-socket peptide optimization. Once having defined a model for the PDZ domain, the next step is to use the knob-socket model to optimize the binding peptide. The knob-socket model utilizes data found in the PDB to determine which residues would fit best in the domain and considers multiple factors. For example, when determining the best amino acid for position P_0 in the peptide, there is an interaction with both Coil2 and Strand2 of the domain's sheet. Figure 5 separates out each of these interactions into the four categories shown and determines the best peptide residues based on position and propensity. The logos plots shown for each category represents the best peptide residues for the function they hold in that category. This is why each logos plot does not contain a prediction for every peptide residue. The specificity of the peptide

depends on multiple factors, which is why each residue can vary depending on which factor is held as more or less significant when binding occurs. Based on what the knob-socket showed as a difference between a bound and unbound PDZ domain, the interaction between the bound peptide and Strand2 of the domain was deemed as most important when ranking the four interactions of Figure 5. As pointed out before, without the peptide's extension of the sheet and the subsequent sockets and pockets formed in this extension, the helix would not be close enough to the domain's sheet to interact with it. This model is able to consider the secondary interactions that would otherwise never be seen. This would then make the second most important interaction between Helix2 of the domain and the newly extended sheet. This interaction, which is a combination of knobs from the helix into the sheet as well as knobs of the peptide into the helix, brings the structure of the PDZ domain together in a way that allows high affinity and also specificity. The remaining interaction, that between the peptide and Coil2 of the domain, is not ranked high due it's interaction with the sheet regardless of the presence of a peptide. In many families of the PDZ domain, it is observed that residues of the unbound domain's own sheet would sometimes pack into the coil, concluding that proximity of the coil to the sheet was not dependent on the presence of a peptide. With all four of these interactions ranked and logos plots indicating "better" residues for the bound peptide for each of those interactions made, a consensus sequence could be created in an attempt to optimize the specificity and affinity of a bound peptide with the PDZ domain.

Consensus peptide sequence. Figure 5 breaks down the individual interactions that are most commonly seen between a bound peptide and the PDZ domain. The next step is to take each of these interactions and bring them together for a common consensus sequence, as seen in Figure 6.

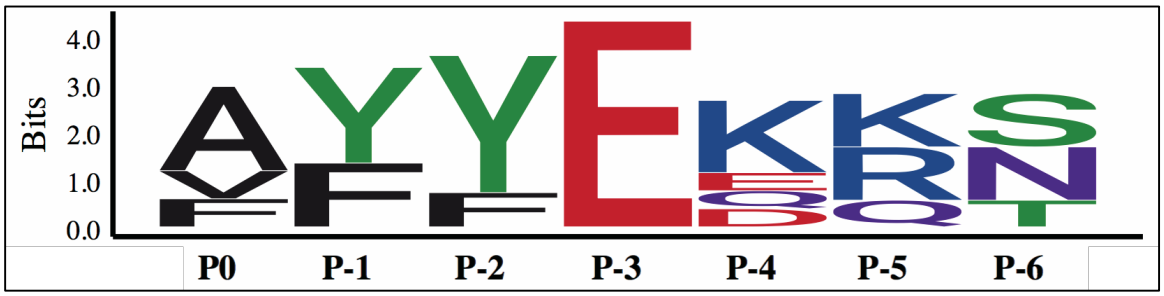


Figure 6: Bound Peptide Consensus Sequence. Description. For each of the four interactions discussed, a consensus sequence was found that satisfied the highest propensities for a bound peptide into a PDZ domain of the Tiam1 family.

This consensus sequence shows the residues with the highest propensity to bind into the PDZ domain by taking into account all four interactions plotted in Figure 5. Ideally, this sequence would represent a peptide with the highest affinity and specificity for the Tiam1 family PDZ domain. The next steps that would need to be taken to verify this statement would be binding studies. While the logos consensus sequence plot in

Figure 6 provides many varieties of a sequence for the bound peptide, it is equally as important to test the binding of peptides with sequences from the four interactions isolated in Figure 5. These calculated interactions may not form a continuous sequence the way the consensus sequence does, but they do present important propensity results that individually account for the strength of the bound peptide. A combination of binding study results from both the individual interaction plots as well as the overall consensus sequence in comparison to that of the naturally found syndecan peptides will help verify the prediction capabilities of the knob-socket model and therefore help bring protein prediction capabilities that much closer to success.

Conclusion

Through the use of the knob-socket model, not only was it possible to visualize the binding interactions between the PDZ domain of the Tiam1 family and a bound peptide on a two-dimensional map, but the common interactions among different domains could be identified and characterized. The five common interactions were found between the bound peptide with Strand2, Coil2, and Helix2 of the domain. The most important of these interactions was determined to be between the bound peptide and Strand2, which acts as an extension of the domain's β-sheet. Not only did the bound peptide produce new sockets between Strand2 and itself, but it also created a binding surface for the residues of Helix2 to pack into, further defining the quaternary structure of the bound PDZ domain. These individual interactions were then further analyzed to

optimize the sequence of the bound peptide into the domain. Figure 5 shows the results from individually optimizing each interaction. The resulting logos plots show that different residues will "fit" for the same position in the bound peptide sequence. For this reason, an overall consensus sequence was plotted, shown in Figure 6, where each of the interactions were brought together to form a continuous sequence that would bind the domain best. The next steps for this work would involve binding studies comparing all of the optimized sequences in this paper to the binding of naturally found peptides for each PDZ domain of varying families. With this comparison, not only will the prediction capabilities of the knob-socket model be tested, but the scientific community would be that much closer to better understand protein folding and binding for globular proteins.

Chapter 3: Analyzing Proteins of the Bcl-2 Domain

One promising approach in selective cancer therapy is to use a cell's own primary immune mechanism of programmed cell death, or apoptosis, against the cancerous state. In a normal cell, apoptosis occurs when a cell detects anomalous conditions. The signal is transmitted through the pro-apoptotic BH3 protein, which then interacts with the multi-domain BAX/BAK complex to trigger programmed cell death (DeBartolo et al., 2012). In healthy cells, BH3 activity is shut off. Cancer cells take advantage of these "off switches" by upregulating the expression of the anti-apoptotic Bcl-2 protein, which binds BH3 and thus prevents any apoptosis signal from moving forward (Dutta et al., 2010).

There are many different groups of proteins that govern whether a cell lives or undergoes apoptosis. Pro-apoptotic and anti-apoptotic proteins of the Bcl-2 domain, in particular, interact in order to determine whether a cell will live or die (Dutta et al., 2010). Specifically, the proteins that share the BH3 domain with Bcl-2 such as BAK and BAX are essential for activating apoptotic processes, as seen in Figure 7. BAX and BAK exist as monomers, but when they receive signals from BH3-only proteins they bind and insert into the mitochondrial membrane, inducing apoptosis (Petros et al., 2004). Cancer cells have several different processes which allow them to disable these processes and help them survive. One of these processes is producing anti-apoptotic Bcl-2 proteins to block pro-apoptotic signals (Muchmore et al., 1996; Petros et al., 2004). Another such

mechanism used by cancer cells is to inhibit apoptosis by mutating the DNA to suppress certain genes. These processes have been organized into three classes: class A, which is the decrease of activator proteins, class B, which is the decrease in BAX and BAK, and class C, which is the increase in inhibitor proteins.

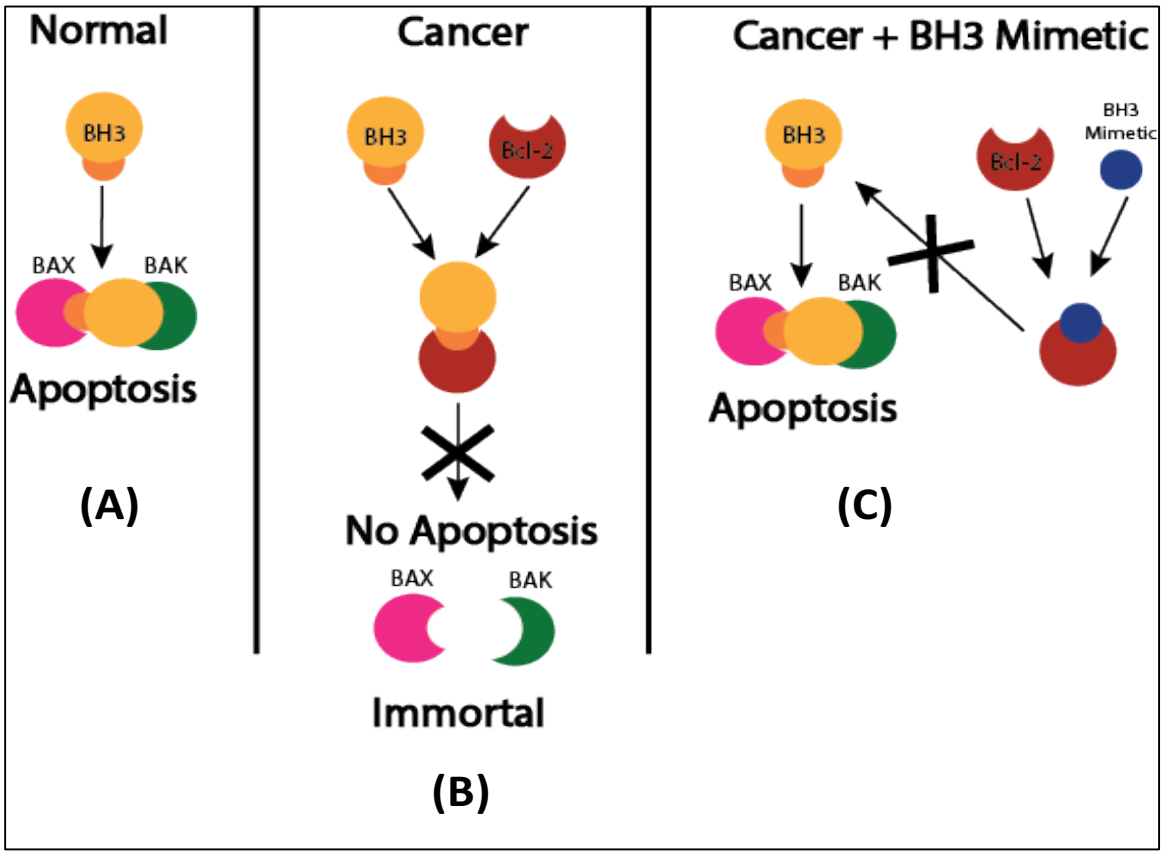


Figure 7: Apoptotic Pathway. (A) A normal cell undergoes apoptosis when pro-apoptotic proteins of the BH3 domain bind with BAX and BAK. (B) A cancerous cell blocks the apoptotic pathway by producing anti-apoptotic proteins of the Bcl-2 domain that bind with BH3 and prevent BAX/BAK activation. (C) BH3 mimetics bind with Bcl-2, freeing the pro-apoptotic BH3 proteins to bind with BAX and BAK and causing apoptosis.

Figure 8 shows the three-dimensional structure for all four of the analyzed complexes in this work. The PDB label above each structure indicates a specific member of the Bcl-2 family binding a helical BH3 ligand. The general structure of members of the Bcl-2 family consists of six to nine helices connected together by random coils (Petros et al., 2004). The agreement amongst those that have studied the structure of Bcl-2 and its ligand is that the protein contains a hydrophobic groove formed by two or three of its helices, and that is where the ligand binds into (Dutta et al., 2010; Muchmore et al., 1996). The BH3 ligand is an α-helical protein that, when bound to members of the Bcl-2 family, halts cellular apoptosis and therefore allows unhealthy cells to survive and multiply without regulation.

Proteins have been developed that can counteract the action of cancer cells and increase apoptosis. It is well understood that BAX and BAK initiation are key, but the best way to cause this activation remains under investigation. Some groups emphasize the importance of activator BH3 proteins, which bind directly to BH3 and activate it and help induce apoptosis (Foight & Keating, 2015; Foight, Ryan, Gulla, Letai, & Keating, 2014). Focusing on the importance of neutralizing anti-apoptotic proteins, other groups attempt to induce apoptosis by creating artificial proteins that bind to them, displacing the activators that they were bound to, and allowing them to bind to BAX and BAK and induce apoptosis (Muchmore et al., 1996).

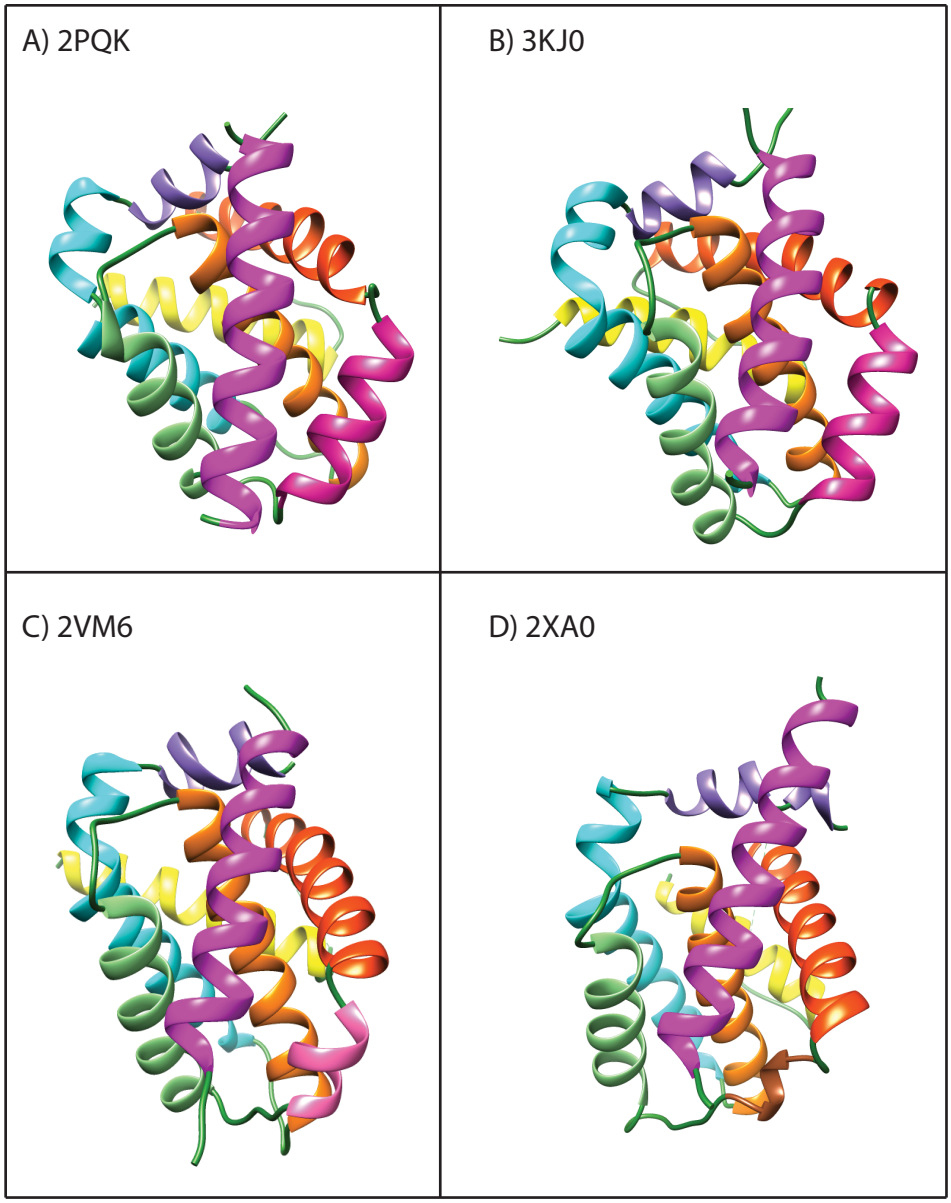


Figure 8: Bound Bcl-2 Family Complexes. Three-dimensional structure of each of the analyzed complexes in this work. The solved structures are from the PDB and displayed using Chimera. The bound ligand has been colored magenta, while each of the helices in the protein is color-coded based on their location in the sequence.

Although certain treatments have been implemented successfully, it is still important to continue research into this topic with the hope of discovering new and more effective treatments. New proteins need to be discovered for the different types of cancers, and the eventual hope is to find a method that can treat many of them. That is why an analysis of the way these proteins pack together is important, so new proteins that could potentially act as more effective treatments can be more easily discovered.

Current Progress in Research

Many groups have attempted to characterize members of the Bcl-2 protein family, which is known to enhance or suppress programmed cell death, of which BH3 proteins are a homologue (Delbridge & Strasser, 2015). By binding an α-helical peptide to a Bcl-2 receptor protein, members of the Bcl-2 family are able to promote cell death. The underlying issue is that homologous pro-death and pro-survival proteins seem to have definitive specificity that has yet to be identified. The determination of these families' specificity and affinity allows for more ways to approach cancer therapy. In particular, more accurate in targeting of cancer cells would preserve more of a person's healthy cells.

Work by Amy Keating's group on this problem has greatly progressed knowledge on the Bcl-2 family function and specificity. In one of her studies, the group looked at interaction specificity from a computational standpoint (DeBartolo et al., 2012). With the use of a position-specific scoring matrix (PSSM) and another statistical potential program

called STATIUM, the group was able to show that computational tools such as these can further aid in the search for Bcl-2 family binding proteins. In another study, the Keating group investigated the Mcl-1 protein, a pro-survival protein (Foight et al., 2014). They chose three peptides, using yeast- surface display, to show binding specificity and affinity to Mcl-1. A peptide was identified, MS1, that was able to bind Mcl-1 at 40-fold higher specificity than any other Bcl-2 homologue. These experiments were performed using methods such as fluorescence anisotropy assays and cellular BH3 profiling assays.

While the work done by Keating's and many other groups has given insight on these Bcl-2 proteins, a true model defining the specificity and affinity of this interaction has yet to be identified. This approach uses the knob-socket motif to shed light on the actual binding patterns present in both the pro-death and pro-survival proteins. By looking at local residue interactions that are occurring in the bound structures, a pattern could be found that distinctly identifies the binding pattern of each protein. With this sort of information, prediction and design of peptides targeting anti- and pro- apoptotic pathway proteins are potentially possible.

Knob-Socket Model Application

By using the knob-socket model to map out the protein-ligand interactions of Bcl and BH3 in the various structures, similarities can be found between the different ligands to identify residues responsible for specificity and to help design effective inhibitors of the anti-apoptotic signal in cancer cells. There are four specific proteins from the Protein

Data Bank (PDB) that were analyzed to look into the interactions between pro- and anti-apoptotic proteins: 2PQK, 3KJO, 2VM6, and 2XAO. A full protein surface topology (PST) map of the protein-ligand interactions involved in the 2PQK protein can be seen in Figure 9. Upon examination of the knobs and sockets that define how these proteins pack with their ligands, it was observed that there was a certain number of conserved residues in the ligand that consistently acted as knobs. A glycine, leucine, and isoleucine in the ligand packed into another helix in all four of the examined proteins. By further analyzing the similarity and specificity of the knobs and sockets in these regions, the goal was to find the defining factors in order to design proteins that are able to have the same interactions.

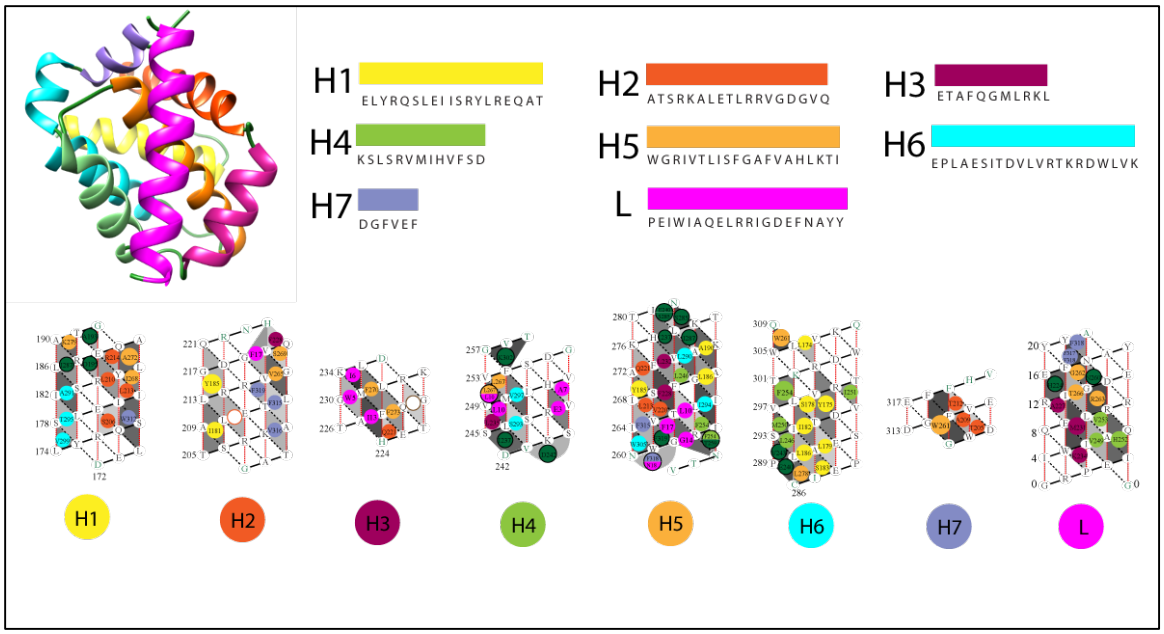


Figure 9: Knob-Socket Mapping of 2PQK. (A) 3D model of the interaction between Mcl-1 and Bim from Chimera. (B) Amino acid sequence of each of the helices. (C) 2D models of the different helices based on the knob-socket model.

Results and Discussion

Through the use of the knob-socket model and the modified helical lattices it utilizes in order to better visualize protein-protein interactions, a few patterns were seen across the observed proteins. There were three amino acids in particular that always packed from the ligand to the protein, which were a glycine, leucine, and isoleucine. These three amino acids always packed into either Helix1, Helix4, Helix6, or Helix8, although the knobs which they packed back into were different for each complex. It is

important to recognize that interactions occurring between the protein and ligand are all helix-helix interactions. This means that the sockets in which ligand knobs are packing are all helical sockets of either XY:H or H:YX residue composition (See Figure 2 for the knob-socket definition of helical interactions). There were also interactions between the ligand and Helix1 in all four of the protein complexes examined. In each of these examples, either a glycine or a leucine knob from the ligand packs into the protein, and glycine, arginine, tyrosine, and tryptophan residues from the protein packs back into the ligand. Figure 10 shows the mapping done for each of the four ligands from the four analyzed protein complexes. The knob colors in this figure correspond to the designated helix number and color seen at the top of Figure 9. For each of the following observations, the glycine knob from the ligand was kept at a zero position so that each of the maps could be compared while remaining in the same frame of reference.

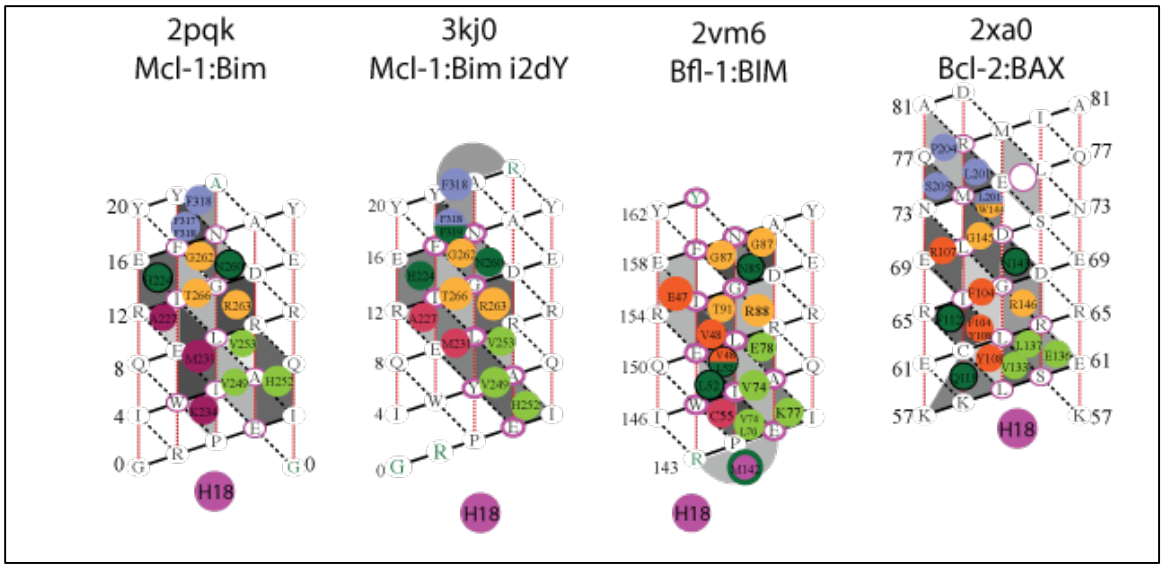


Figure 10: Helical Ligand Maps. Two-dimensional models of the ligand helices of all four interactions made using the knob-socket model. Each of the helices is centered around the glycine, isoleucine, and leucine amino acids.

The other helices into which the ligand packs are Helix8, Helix4, and Helix6, but these were not consistent interactions across every complex. Helix6 and Helix8 have interactions in both 2PQK and 3KJ0. An isoleucine in the ligand packs into a socket comprised of histidine, alanine, phenylalanine, and methionine. The alanine from that socket in the protein packs back into a socket in the ligand of glutamine, arginine, and isoleucine. Likewise, the methionine packs back into a socket of tyrosine, glutamine, isoleucine, and leucine as well. 3KJ0 and 2PQK have very similar binding patterns because both of these complexes include Mcl-1 and some form of BH3.

Helix6 interacts with the ligand in the proteins 2XA0 and 2VM6. In 2XA0, an isoleucine from the ligand packs into a large socket of two phenylalanine residues, two arginine residues, a tyrosine, and a glutamic acid in the protein. The arginine from the protein packs back into a socket in the ligand comprised of an isoleucine, a glutamine, a leucine, and an asparagine. A phenylalanine from the protein also packs back into a ligand socket of two leucine residues, an isoleucine, and a glycine. Finally, a tyrosine of the protein packs back into a ligand socket of two leucine residues, a cysteine, and an isoleucine.

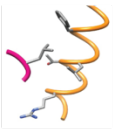
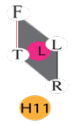
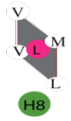
In 2VM6, the isoleucine from the ligand instead packs into a socket of serine, two valine residues, and a glutamine. The glutamine from the protein packs back into a ligand socket of two glutamine residues, an arginine, an isoleucine, and a phenylalanine. The protein's valine knob also packs into a ligand socket of two isoleucine residues, a glutamine, and a leucine. Lastly, a cysteine packs back into a ligand socket of proline, tryptophan, and isoleucine.

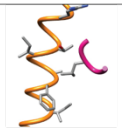
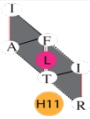
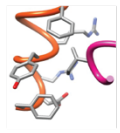
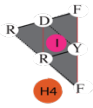
These observations of interactions between the ligand and Helix8, Helix4, and Helix6 were not observed as consistent among all four of the mapped protein complexes, but they still provided valuable information that was easily found through the use of the knob-socket model and its two-dimensional modified lattice mapping.

Table 1 breaks down each interaction that was observed across each of the mapped protein-ligand complexes. These interactions specifically look at the three main

amino acids from the ligand that are conserved: glycine, leucine, and isoleucine. The binding of the ligand to the protein was shifted according to the location of these three residues in the ligand helix. These three residues contributed consistently to a hydrophobic groove that is key to the interaction between the activator and receptor proteins. The future work for this project involves collating these maps into models of helix binding for pro-survival or pro-apoptotic binding. The use of these results in creating a BH3 mimetic could have potential use as a treatment.

Table 1: Isolated Protein-Ligand Interactions. Key interactions involving isoleucine, leucine, and glycine from the ligand. Each interaction is systematically named, with a picture of the three-dimensional model for the specific interaction and the two-dimensional model of the interaction based on the knob-socket model.

INTERACTION	PICTURE	KNOB-SOCKET
H11.HP.G1 : Glycine Packing: 2PQK, 2VM6, 3KJ0		
H11.HP.G2: Glycine Packing: 2XA0		
H11.HP.L1: Leucine Packing: 2PQK, 3KJ0		
H8.HP.L1: Leucine Packing: 2PQK, 3KJ0		
H11.HP.L2: Leucine Packing: 2XA0		

INTERACTION	PICTURE	KNOB-SOCKET
H11.HP.L3: Leucine Packing: 2VM6		
H6.HP.I1: Isoleucine Packing: 2PQK, 3KJ0		
H4.HP.I2: Isoleucine Packing: 2XA0		
H4.HP.I3: Isoleucine Packing: 2VM6		

Chapter 4: Conclusion

The knob-socket construct has provided a framework and a language to distinctly characterize the packing detail for both the PDZ domain and the Bcl-BH3 complex. While many research groups have been able to identify some of the interaction sites that occur in both of these systems and their respective binding ligands, they have been limited to thinking only about each residue in the ligand. The knob-socket model can isolate specific tertiary and quaternary interactions in order to better understand the affinity and specificity in each complex. Furthermore, the knob-socket model identified residue inter-dependencies that are necessary for binding. For the PDZ domain, the mapping of the protein and its ligand using the knob-socket data showed an extension of the domain's sheet with the bound ligand as well as tertiary packing of the domain's helix into the newly extend sheet. This produces a tighter packing structure since the sheet-helix interaction did not occur in unbound PDZ domains. For the Bcl-2 complex, a pattern was observed involved the bound ligand's glycine, leucine, and isoleucine residues. These three residues were consistently seen as packing into the protein. Even if those three residues were located in a different region of the ligand, the ligand would shift to allow the packing of those three residues in the same manner. The knob-socket model is essential in identifying these patterns, especially since its mapping method helps break-down and isolate what would have otherwise been a complicated three-dimensional quaternary- level packed globular protein. Patterns such as these help progress the

understanding of protein packing and when supplemented with experimental confirmation of the binding patterns, could be the next steps in finding a safer cure for diseases such as cancer.

REFERENCES

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.

Cheng, B., Montmasson, M., Terradot, L., & Rousselle, P. (2016). Syndecans as Cell Surface Receptors in Cancer Biology. A Focus on their Interaction with PDZ Domain Proteins. *Front Pharmacol, 7*, 10. doi:10.3389/fphar.2016.00010

DeBartolo, J., Dutta, S., Reich, L., & Keating, A. E. (2012). Predictive Bcl-2 family binding models rooted in experiment or structure. *J Mol Biol, 422*(1), 124-144. doi:10.1016/j.jmb.2012.05.022

Delbridge, A. R., & Strasser, A. (2015). The BCL-2 protein family, BH3-mimetics and cancer therapy. *Cell Death Differ, 22*(7), 1071-1080. doi:10.1038/cdd.2015.50

Dutour Sikiric, M., Garber, A., Schurmann, A., & Waldmann, C. (2016). The complete classification of five-dimensional Dirichlet-Voronoi polyhedra of translational lattices. *Acta Crystallogr A Found Adv, 72*(Pt 6), 673-683. doi:10.1107/S2053273316011682

Dutta, S., Gulla, S., Chen, T. S., Fire, E., Grant, R. A., & Keating, A. E. (2010). Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *J Mol Biol, 398*(5), 747-762. doi:10.1016/j.jmb.2010.03.058

Ernst, A., Appleton, B. A., Ivarsson, Y., Zhang, Y., Gfeller, D., Wiesmann, C., & Sidhu, S. S. (2014). A structural portrait of the PDZ domain family. *J Mol Biol, 426*(21), 3509-3519. doi:10.1016/j.jmb.2014.08.012

Fire, E., Gulla, S. V., Grant, R. A., & Keating, A. E. (2010). Mcl-1-Bim complexes accommodate surprising point mutations via minor structural changes. *Protein Sci, 19*(3), 507-519. doi:10.1002/pro.329

Foight, G. W., & Keating, A. E. (2015). Locating Herpesvirus Bcl-2 Homologs in the Specificity Landscape of Anti-Apoptotic Bcl-2 Proteins. *J Mol Biol, 427*(15), 2468-2490. doi:10.1016/j.jmb.2015.05.015

Foight, G. W., Ryan, J. A., Gulla, S. V., Letai, A., & Keating, A. E. (2014). Designed BH3 peptides with high affinity and specificity for targeting Mcl-1 in cells. *ACS Chem Biol, 9*(9), 1962-1968. doi:10.1021/cb500340w

Fuentes, E. J., Der, C. J., & Lee, A. L. (2004). Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J Mol Biol, 335*(4), 1105-1115.

Herman, M. D., Nyman, T., Welin, M., Lehtio, L., Flodin, S., Tresaugues, L., . . . Nordlund, P. (2008). Completing the family portrait of the anti-apoptotic Bcl-2 proteins: crystal structure of human Bfl-1 in complex with Bim. *FEBS Lett, 582*(25-26), 3590-3594. doi:10.1016/j.febslet.2008.09.028

Joo, H., Chavan, A. G., Phan, J., Day, R., & Tsai, J. (2012). An amino acid packing code for alpha-helical structure and protein design. *J Mol Biol, 419*(3-4), 234-254. doi:10.1016/j.jmb.2012.03.004

Ku, B., Liang, C., Jung, J. U., & Oh, B. H. (2011). Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX. *Cell Res, 21*(4), 627-641. doi:10.1038/cr.2010.149

Kundu, K., & Backofen, R. (2014). Cluster based prediction of PDZ-peptide interactions. *BMC Genomics, 15 Suppl 1*, S5. doi:10.1186/1471-2164-15-S1-S5

Lee, H. J., & Zheng, J. J. (2010). PDZ domains and their binding partners: structure, specificity, and modification. *Cell Commun Signal, 8*, 8. doi:10.1186/1478-811X-8-8

Lee, S. O., Lee, M. K., Ku, B., Bae, K. H., Lee, S. C., Lim, H. M., . . . Chi, S. W. (2016). High-resolution crystal structure of the PDZ1 domain of human protein tyrosine phosphatase PTP-Bas. *Biochem Biophys Res Commun, 478*(3), 1205-1210. doi:10.1016/j.bbrc.2016.08.095

Liu, X., Shepherd, T. R., Murray, A. M., Xu, Z., & Fuentes, E. J. (2013). The structure of the Tiam1 PDZ domain/ phospho-syndecan1 complex reveals a ligand conformation that modulates protein dynamics. *Structure, 21*(3), 342-354. doi:10.1016/j.str.2013.01.004

Liu, X., Speckhard, D. C., Shepherd, T. R., Sun, Y. J., Hengel, S. R., Yu, L., . . . Fuentes, E. J. (2016). Distinct Roles for Conformational Dynamics in Protein-Ligand Interactions. *Structure, 24*(12), 2053-2066. doi:10.1016/j.str.2016.08.019

Mignon, D., Panel, N., Chen, X., Fuentes, E. J., & Simonson, T. (2017). Computational Design of the Tiam1 PDZ Domain and Its Ligand Binding. *J Chem Theory Comput, 13*(5), 2271-2289. doi:10.1021/acs.jctc.6b01255

Muchmore, S. W., Sattler, M., Liang, H., Meadows, R. P., Harlan, J. E., Yoon, H. S., . . . Fesik, S. W. (1996). X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature, 381*(6580), 335-341. doi:10.1038/381335a0

Petros, A. M., Olejniczak, E. T., & Fesik, S. W. (2004). Structural biology of the Bcl-2 family of proteins. *Biochim Biophys Acta, 1644*(2-3), 83-94. doi:10.1016/j.bbamcr.2003.08.012

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem, 25*(13), 1605-1612. doi:10.1002/jcc.20084

Qin, X. R., Hayashi, F., Yokoyama, S. (To be published). Solution structure of the PDZ domain of T-cell lymphoma invasion and metastasis 1 varian.

Shepherd, T. R., & Fuentes, E. J. (2011). Structural and thermodynamic analysis of PDZ-ligand interactions. *Methods Enzymol, 488*, 81-100. doi:10.1016/B978-0-12-381268-1.00004-5

Shepherd, T. R., Hard, R. L., Murray, A. M., Pei, D., & Fuentes, E. J. (2011). Distinct ligand specificity of the Tiam1 and Tiam2 PDZ domains. *Biochemistry, 50*(8), 1296-1308. doi:10.1021/bi1013613

Shepherd, T. R., Klaus, S. M., Liu, X., Ramaswamy, S., DeMali, K. A., & Fuentes, E. J. (2010). The Tiam1 PDZ domain couples to Syndecan1 and promotes cell-matrix adhesion. *J Mol Biol, 398*(5), 730-746. doi:10.1016/j.jmb.2010.03.047

Smith, C. A., & Kortemme, T. (2010). Structure-based prediction of the peptide sequence space recognized by natural and synthetic PDZ domains. *J Mol Biol, 402*(2), 460-474. doi:10.1016/j.jmb.2010.07.032

Smith, C. A., Shi, C. A., Chroust, M. K., Bliska, T. E., Kelly, M. J. S., Jacobson, M. P., & Kortemme, T. (2013). Design of a phosphorylatable PDZ domain with peptide-specific affinity changes. *Structure, 21*(1), 54-64. doi:10.1016/j.str.2012.10.007

Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., . . . Sidhu, S. S. (2008). A specificity map for the PDZ domain family. *PLoS Biol, 6*(9), e239. doi:10.1371/journal.pbio.0060239

APPENDIX A. SUPPLEMENTAL DATA FOR PDZ DOMAIN MAPPING

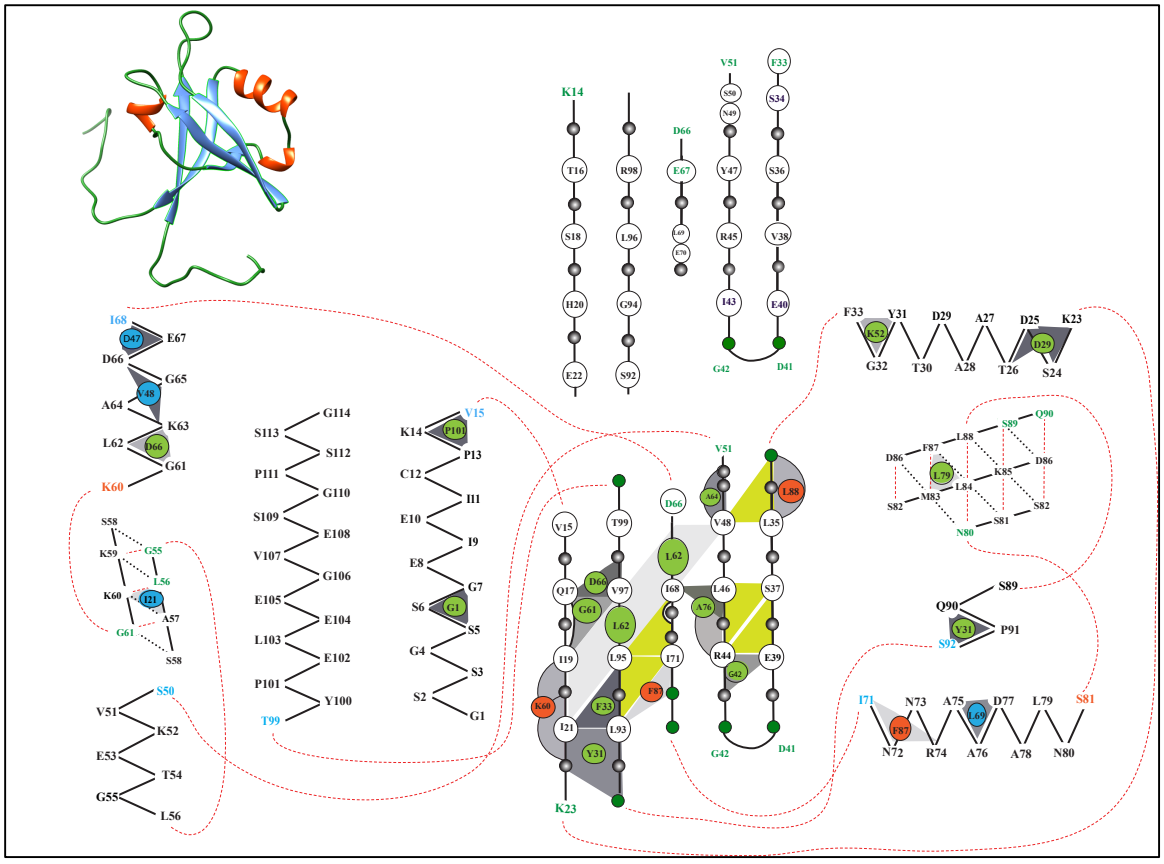


Figure 11: Knob-Socket Mapping of Unbound Tiam1 PDZ Domain. PDB ID: 2D8I (Qin, To be published).

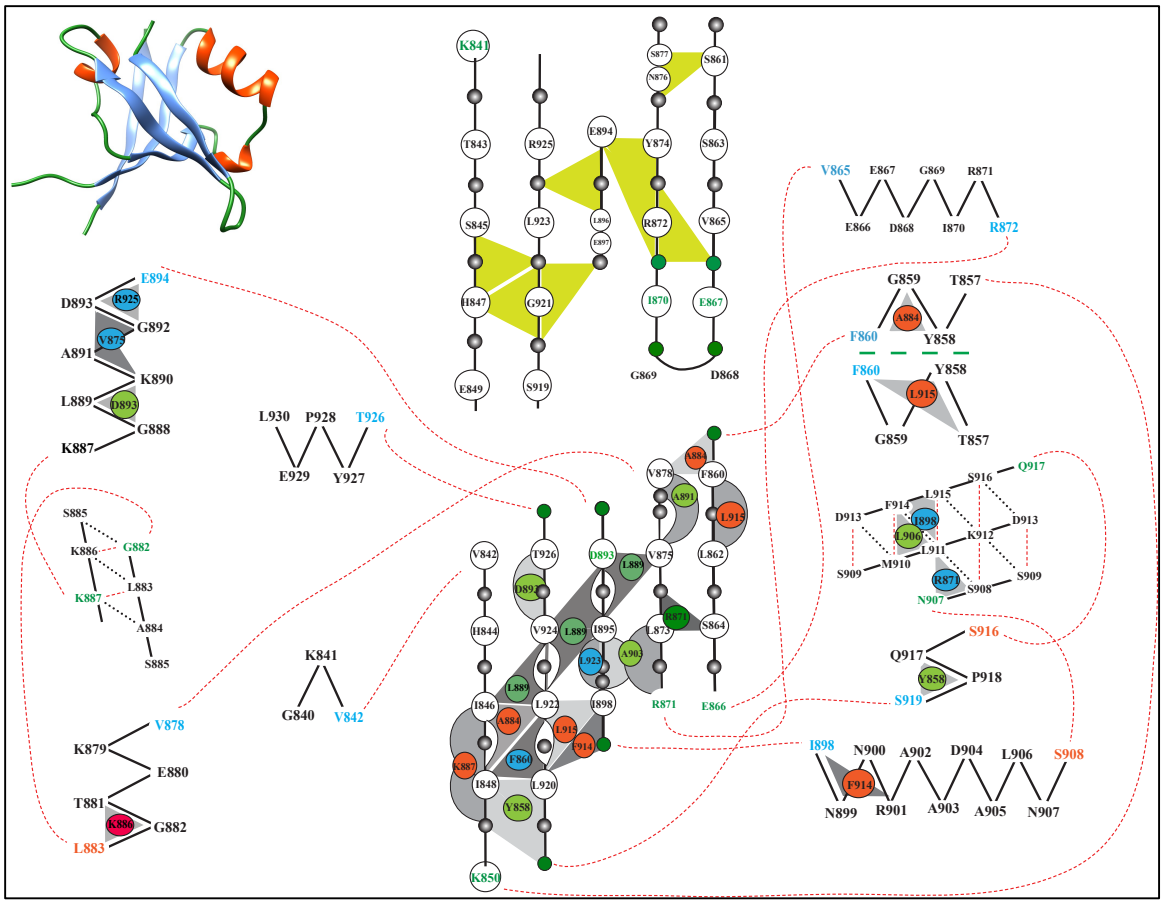


Figure 12: Knob-Socket Mapping of Unbound Tiam1 PDZ Domain. PDB ID: 3KZD (Shepherd et al., 2010).

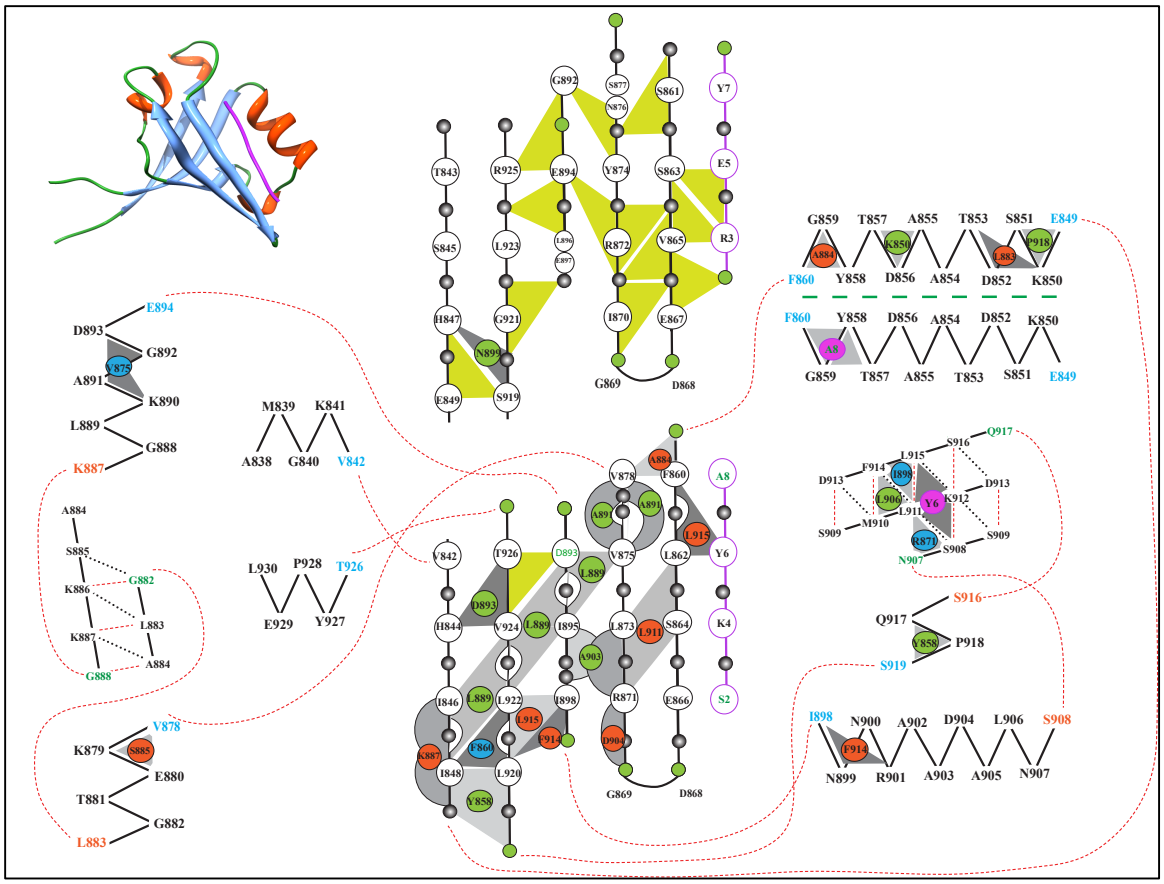


Figure 13: Knob-Socket Mapping of Bound Tiam1 PDZ Domain. PDB ID: 3KZE_BD (Shepherd et al., 2010).

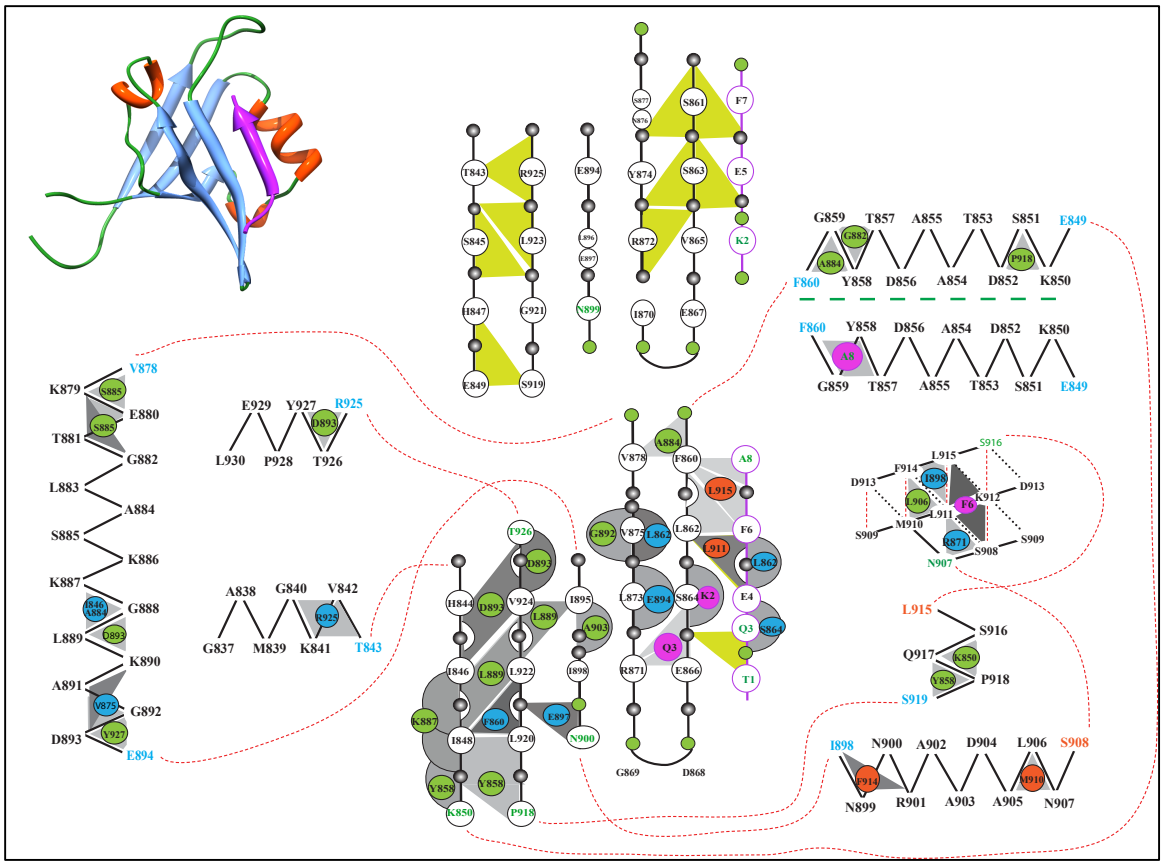


Figure 14: Knob-Socket Mapping of Bound Tiam1 PDZ Domain. PDB ID: 4GVC_AB (Liu et al., 2013).

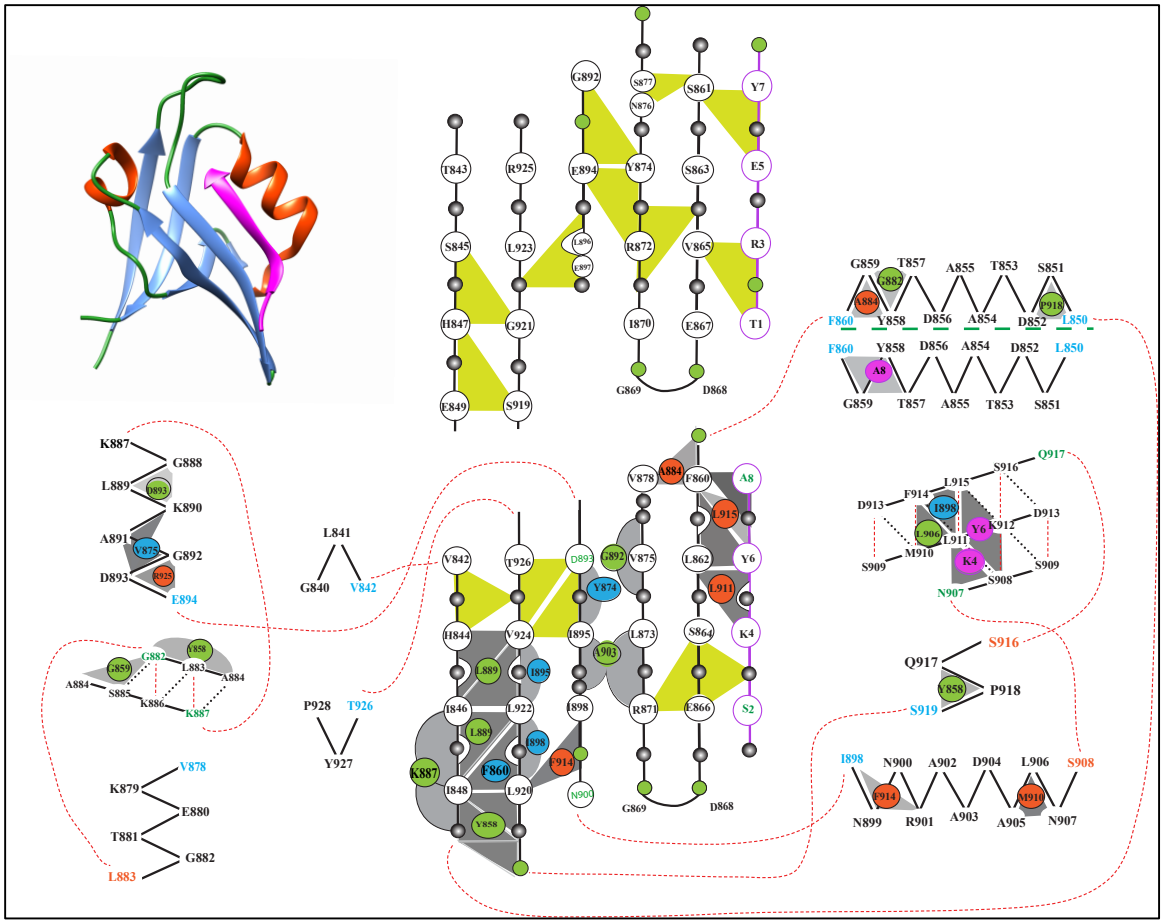


Figure 15: Knob-Socket Mapping of Bound Tiam1 PDZ Domain. PDB ID: 4GVD_BC (Liu et al., 2013).

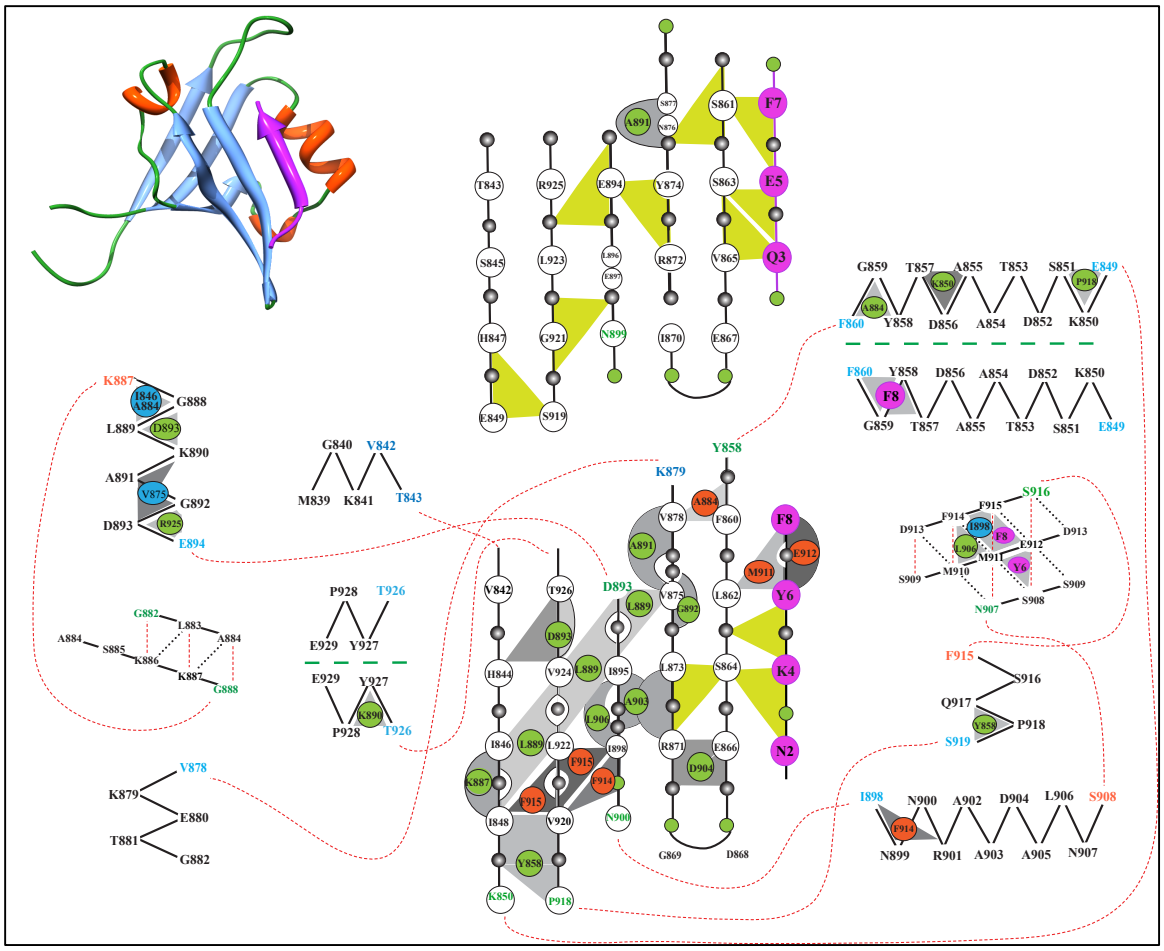


Figure 16: Knob-Socket Mapping of Bound Tiam1 PDZ Domain. PDB ID: 4NXQ_AD (Liu et al., 2016).

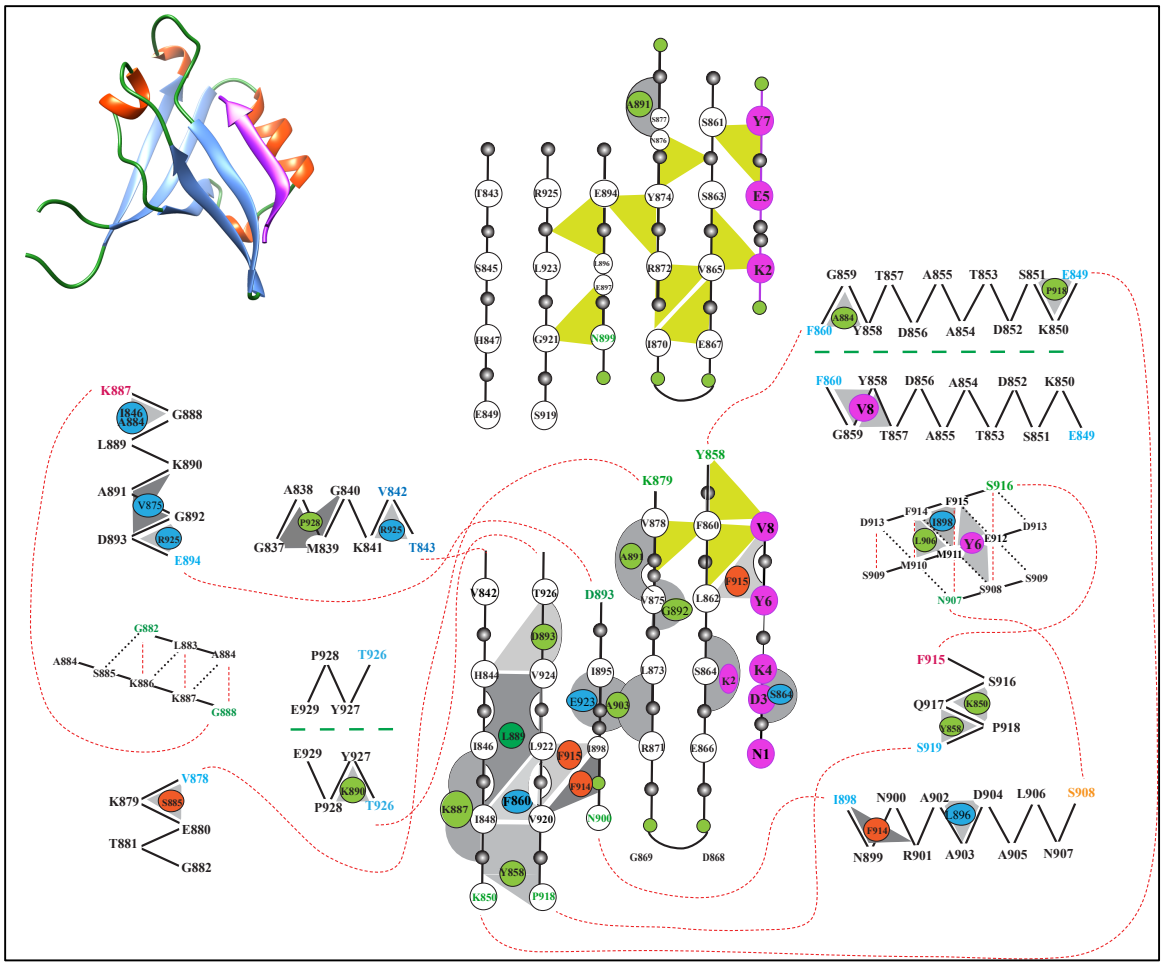


Figure 17: Knob-Socket Mapping of Bound Tiam1 PDZ Domain. PDB ID: 4NXR_AB (Liu et al., 2016).

APPENDIX B. SUPPLEMENTAL DATA FOR Bcl-2 MAPPING

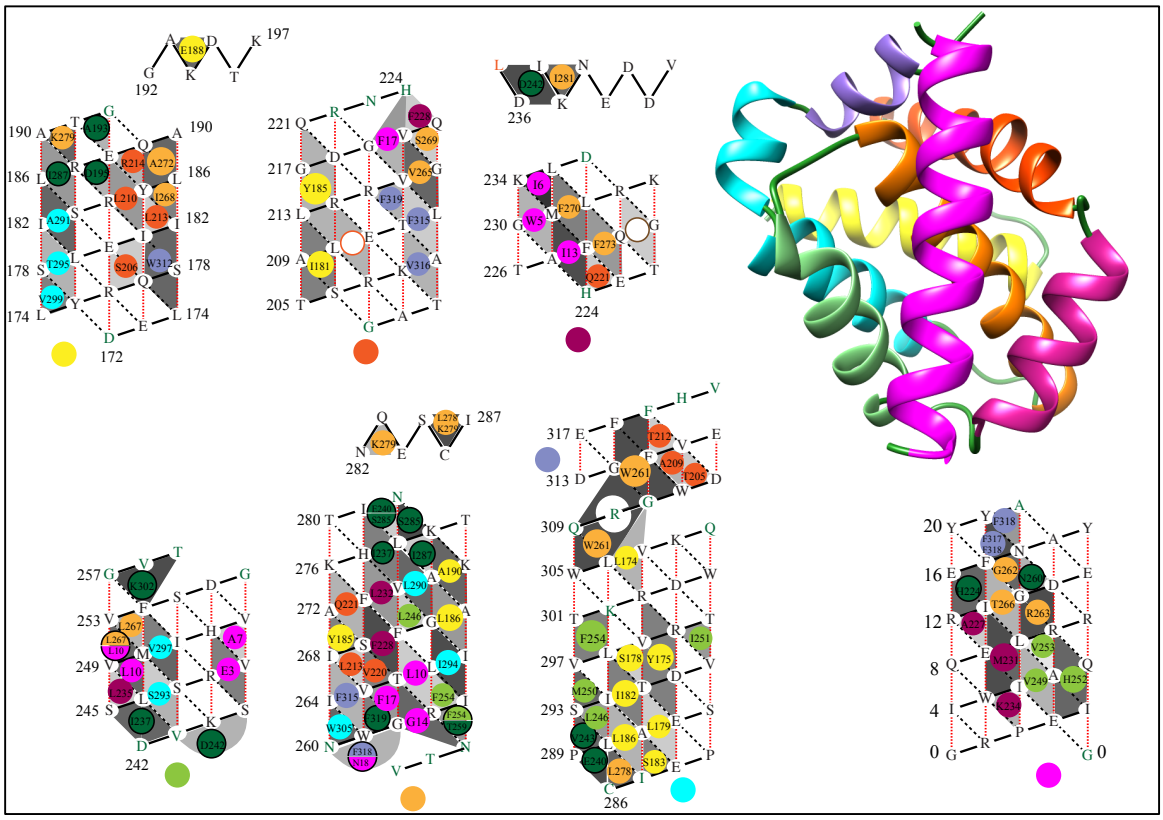


Figure 18: Knob-Socket Mapping of Mcl-1/Bim BH3 Complex. PDB ID: 2PQK (Fire, Gulla, Grant, & Keating, 2010)

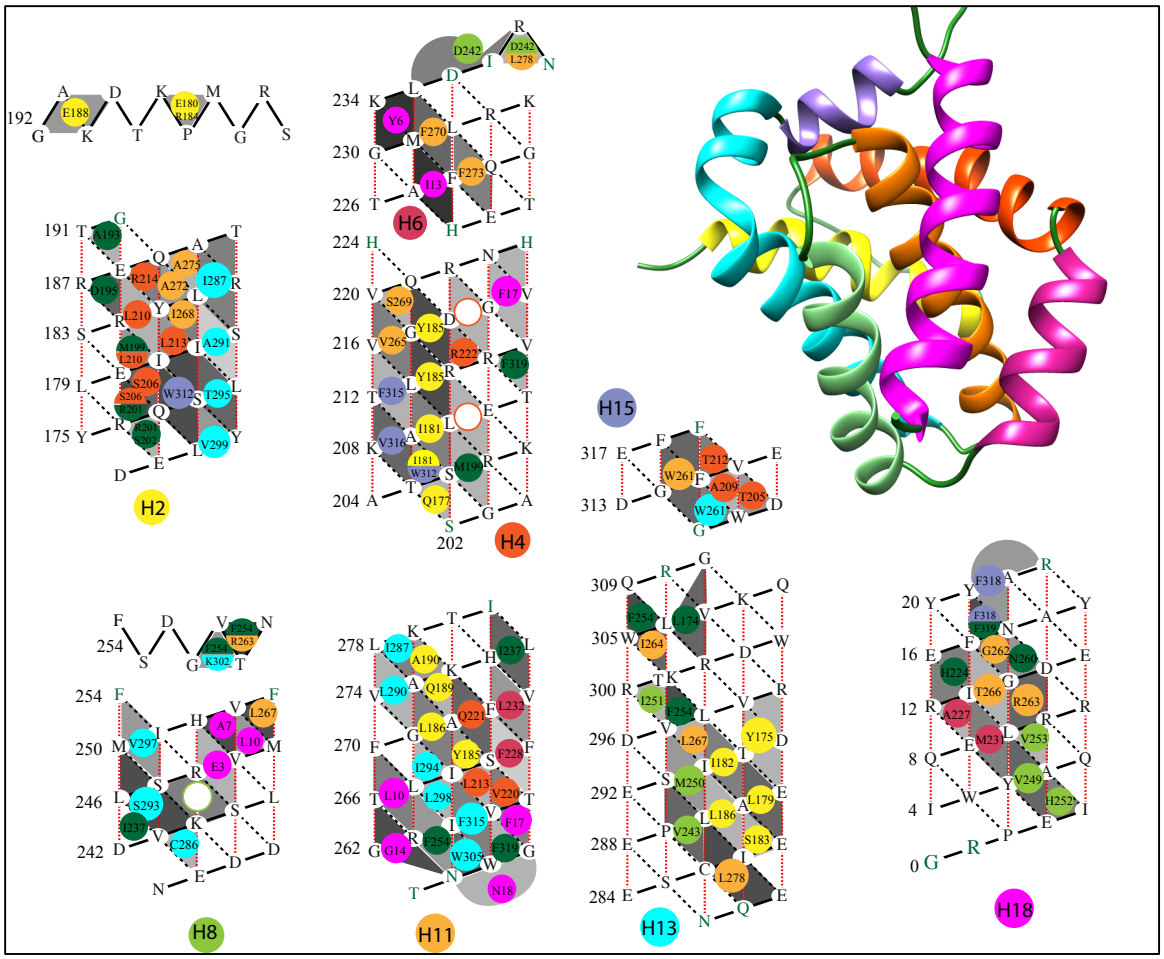


Figure 19: Knob-Socket Mapping of Mcl-1/Bim BH3 Complex. PDB ID: 3KJ0 (Fire et al., 2010)

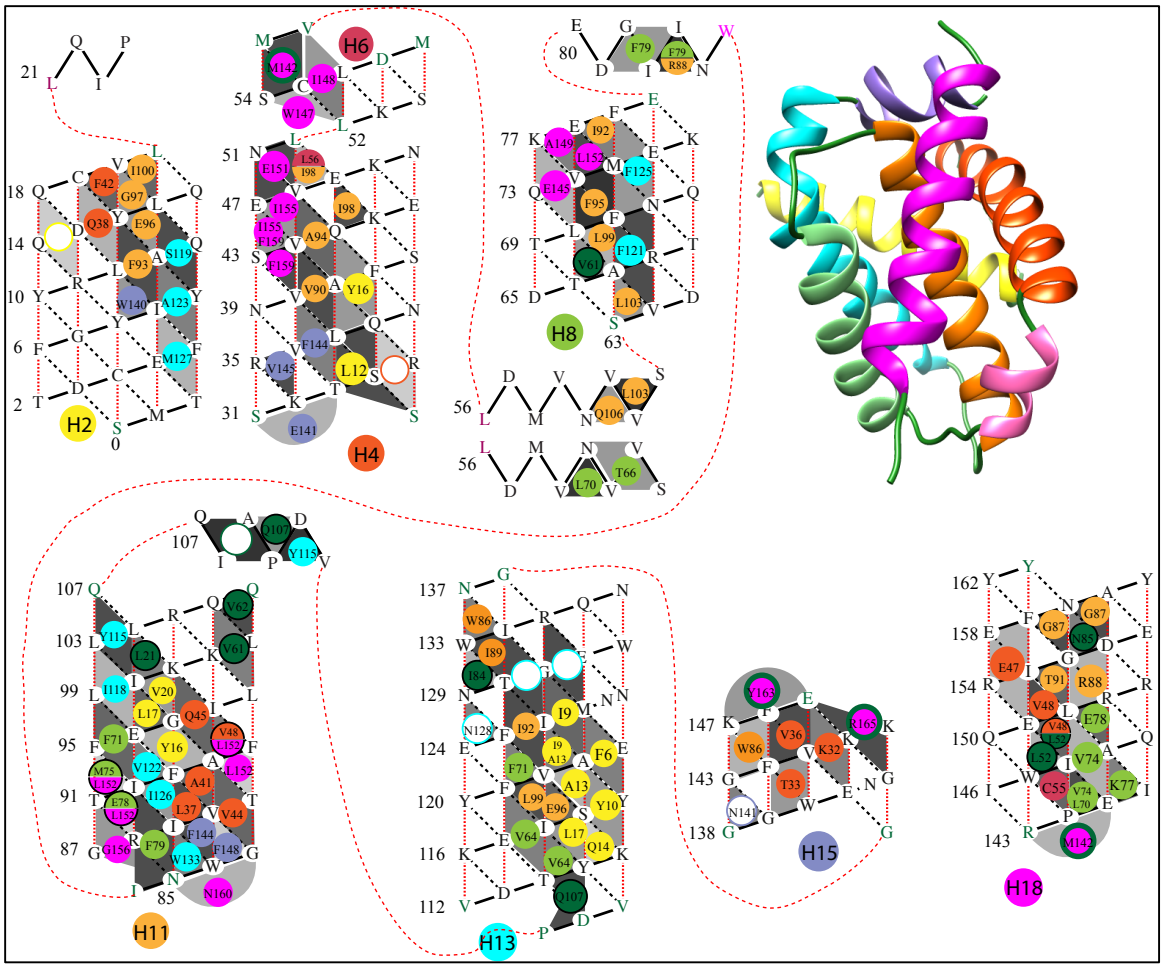


Figure 20: Knob-Socket Mapping of Bcl-2/Bim-BH3 Complex. PDB ID: 2VM6 (Herman et al., 2008)

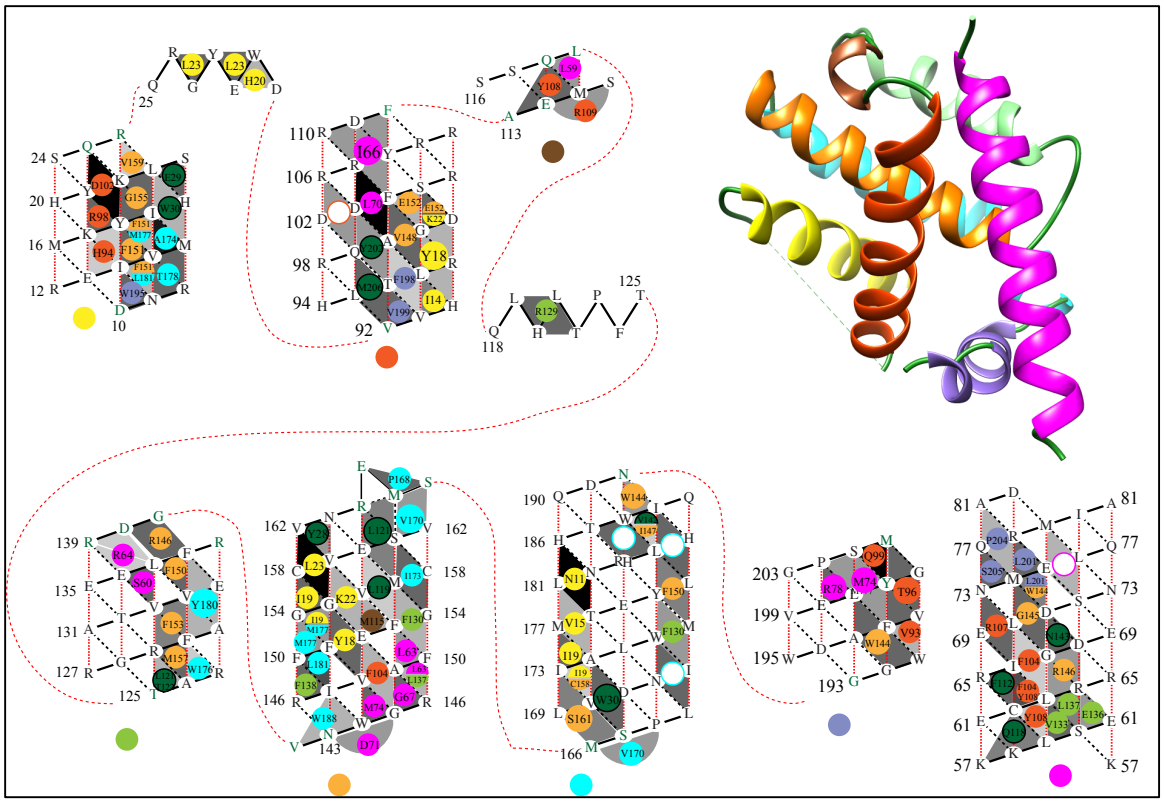


Figure 21: Knob-Socket Mapping of Bcl-2/BAX BH3 Complex. PDB ID: 2XA0 (Ku, Liang, Jung, & Oh, 2011)